U.S. Department of Justice
Federal Bureau of Investigation

PROCEEDINGS OF THE INTERNATIONAL SYMPOSIUM ON THE FORENSIC ASPECTS OF DNA ANALYSIS
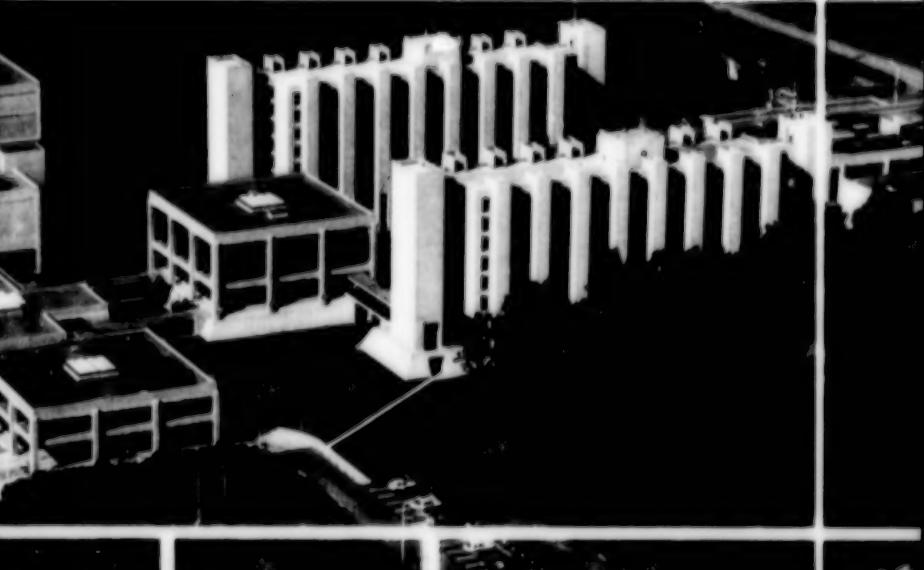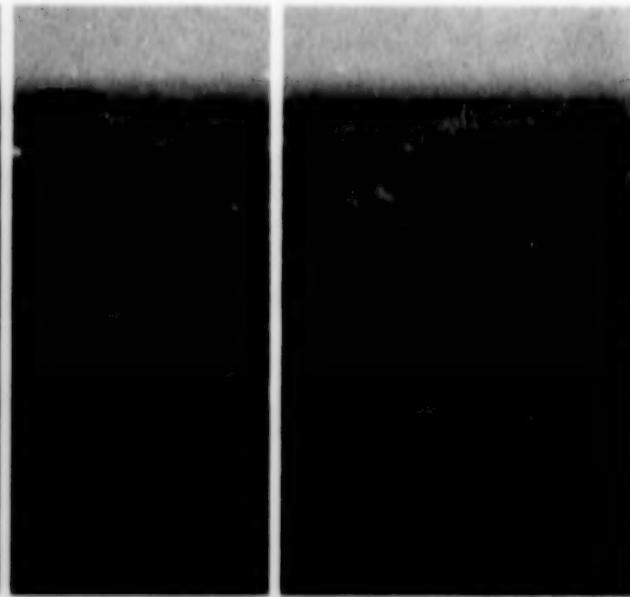
FBI ACADEMY
QUANTICO, VIRGINIA
JUNE 19-23, 1989

Proceedings

of the

International Symposium

on the

Forensic Aspects of DNA Analysis

Host
Laboratory Division
Federal Bureau of Investigation

June 19-23, 1989

Forensic Science Research and Training Center

FBI Academy

Quantico, Virginia

NOTICE

This publication was prepared by the U.S. Government. Neither the U.S. Government nor the U.S. Department of Justice nor any of their employees makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product or process disclosed, or represents that in use would not infringe privately owned rights. Reference herein to any specific commercial product, process or service by trade name, mark, manufacturer or otherwise does not necessarily constitute or imply its endorsement, recommendation or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

Cover: Aerial photograph of the FBI Academy by George February.

FOREWORD

On June 19-23, 1989, the FBI Laboratory hosted an "International Symposium on the Forensic Aspects of DNA Analysis" at the FBI Academy, Quantico, Virginia. This symposium was attended by 366 scientists from industry, academia and forensic science laboratories from 20 foreign countries and nearly every state in the United States.

The law enforcement community must utilize all available resources in the solution of violent crimes. For the crime laboratory, this means determining as much information as possible about the criminal from the physical evidence left at the scene. Recently there has been increased emphasis placed on the examination of biological evidence due to developments in the area of DNA analysis.

Obviously, the ultimate goal for the forensic scientist when examining biological evidence would be the identification of an individual. This has become possible through the analysis of genetic material, DNA. The FBI initiated an aggressive research program to develop DNA analysis technology for implementation into the FBI Laboratory. This research also sought to establish the scientific validity and reliability of these methods to insure that evidence derived from the forensic analysis of DNA could withstand legal challenges inherent with its introduction into the courts. Then the FBI initiated an ambitious technical training program to transfer the knowledge of this rapidly evolving technology to the state and local crime laboratories for their use.

The FBI research effort was directed through the FBI Laboratory's Forensic Science Research and Training Center at the FBI Academy, Quantico, Virginia. Simultaneous research was conducted on the potential forensic analytical methods of Restriction Fragment Length Polymorphism, Polymerase Chain Reaction and DNA sequencing. The research effort was augmented through collaboration with the crime laboratory community by their participation in the FBI Laboratory's Visiting Science Program. Due to the technical nature of the DNA analysis procedures, it was necessary to develop a 4-week course just to provide the basic theory and some hands-on training. In addition, laboratory and classroom resources were reallocated to accommodate the 30 students per class. Currently, 120 students per year are receiving this training.

To further insure the DNA analysis technology methods developed for implementation into the crime laboratory community and the judicial system would be accepted, additional cooperation with the forensic community was sought. This led the FBI Laboratory to host several meetings of a working group of experts in forensic serology who had provided a leadership role in the national implementation of the rapidly developing DNA analytical capabilities. This working group recognized that in order to obtain the newest and best information relating to DNA analysis, they would have to gather scientists from industry and academia as well as forensic laboratories. With input from the working group and others, a small forensic science seminar on DNA technology was hosted in 1988. Due to the continued development in DNA technology worldwide, this present international symposium was planned.

The exchange of ideas at this symposium will undoubtedly generate future research interest into DNA analysis and result in a strengthening of the scientific merit of these examinations. I think that by working together, there is much we can accomplish. I can assure you that the FBI is committed to the continuance of its strong pursuit toward the further development of the science in the field of forensic applications of DNA.

On behalf of the FBI, I would like to thank all those who participated in making this symposium a success.

William S. Sessions
Director
Federal Bureau of Investigation

iii

iv

Program Organizing Committee

Dwight Adams
FBI Laboratory

F. Samuel Baechtel
FBI Laboratory

Bobby D. Blackburn
FBI Laboratory

Barry L. Brown
FBI Laboratory

Bruce Budowle
FBI Laboratory

George W. Clarke
*San Diego Office of the District
    Attorney*

James J. Kearney
FBI Laboratory

Lawrence A. Presley
FBI Laboratory

Willard Carl Stuver
*Metro-Dade Crime Laboratory
    Bureau*

Richard Tanton
*Palm Beach Sheriff's Department
    Crime Laboratory*

Session Moderators

Bobby D. Blackburn
FBI Laboratory

Bruce Budowle
FBI Laboratory

John W. Hicks
FBI Laboratory

James J. Kearney
FBI Laboratory

Pamela Newell
Centre of Forensic Sciences

Lawrence A. Presley
FBI Laboratory

Richard Tanton
*Palm Beach Sheriff's Department
    Crime Laboratory*

John S. Waye
Royal Canadian Mounted Police

David J. Werrett
Home Office Forensic Science Service

BLANK PAGE

CONTENTS

FOREWORD

SECTION I—LECTURES

SECTION III—PANEL DISCUSSIONS

SECTION I

LECTURES

BLANK PAGE

DNA: THE HISTORY AND FUTURE USE IN FORENSIC ANALYSIS

C. Thomas Caskey, Al Edwards and Holly A. Hammond

Baylor College of Medicine
Houston, Texas

The fields of molecular and human genetics provide new and improving DNA based identification methods for law enforcement via the forensic science field. The application of DNA profiling will be held to the highest standards, since its identifying capacity can bring convictions previously unattainable and acquittal previously unsuspected. Defense law firms and their expert witnesses will attach every aspect of the science and methods in their attempts to discredit their validity when it serves the interest of their client, and they will embrace the technology when it serves to their client's advantage. Defense attorneys and their expert witnesses may make gains occasionally as reflected by recent articles in the New York Times, the Wall Street Journal, and an editorial in Nature (Lander 1989). Judges may not admit the evidence in some cases.[1] These gains by defense expert witnesses will very likely be limited, since the molecular and human genetic principals are sound and can be applied successfully and precisely to the forensic science field.

In this paper the history of DNA based technological advances in the basic, medical, and forensic sciences is summarized. Standards for forensic laboratories and samples are considered. The use of the variability between the DNA of individuals in the human population in current and future methods for DNA profiling is discussed. In particular, a new method for DNA typing based on the presence of highly polymorphic short tandem repeat or dispersed minisatellite sequences within the human genome is presented.

HISTORY OF MOLECULAR DNA ADVANCEMENTS IN THE BASIC SCIENCES, MEDICINE AND FORENSICS

The brevity of the involvement of DNA based methodologies in the forensic science field can be illustrated by indicating the milestones in the technology that have led to advances in the fields of medicine and forensic science. Figure 1 illustrates the advances in DNA technology over the period of time from the mid-fifties up until present. The sentinel advances initiated with the description of the double helical structure of the DNA molecule. This was followed by the elucidation of the genetic code, the development of molecular cloning techniques, DNA sequencing, the identification of restriction fragment linked polymorphisms (RFLPs) and development of the polymerase chain reaction (PCR). The PCR method is of great importance in the future of forensic science since it enables the accurate and specific copying and amplification of small amounts of specific DNA targets in a test tube. There has been a gradual improvement in DNA methods, punctuated by major advances in technology.

What is the time scale of application of these DNA based technologies to the field of medicine and forensic science? One of the early advances in the medical field brought about via the application of molecular methods was the cloning of the hemoglobin gene. This led rapidly to the diagnosis of sickle-cell disease (Figure 2). The technology has led to the identification, isolation and characterization of oncogenes. The RFLP linkage strategy has been used to map disease genes, and permitted reverse genetic technology for the identification of disease genes, which had proven refractory to classical approaches (Orkin 1986). Figure 2 illustrates the rapid and increasing application of the technology in medicine beginning in the later seventies. The initial DNA diagnostic laboratories for medicine began delivering their first formal reports in 1984 (Caskey and White 1983).

Consider the application of DNA based methods to the forensic field (Figure 3). Alec Jeffreys was first to bring the area to our attention in his sentinel papers on minisatellites in 1985 (Jeffreys *et al.* 1985; Gill *et al.* 1985). So-called "DNA fingerprinting" utilizing these minisatellite sequences or variable number tandem repeats (VNTRs) was first applied to forensic cases in the United Kingdom, and later in the United States, with the admission of the FBI cases this past year. Figure 3 illustrates that the application of DNA based technologies to forensic problems is a very recent event in the his-

[1] Since this presentation Judge Gerald Scheindlin ruled not to admit the DNA data on the basis of his interpretation that the specific studies in the case were not valid on technical grounds. Judge Scheindlin did, however, rule in favor of the general admissibility of DNA forensic data. Castro subsequently confessed to the double murder and admitted that the blood on his watch came from the victims. This was the interpretation of Life Code's DNA analysis not admitted to jury review.
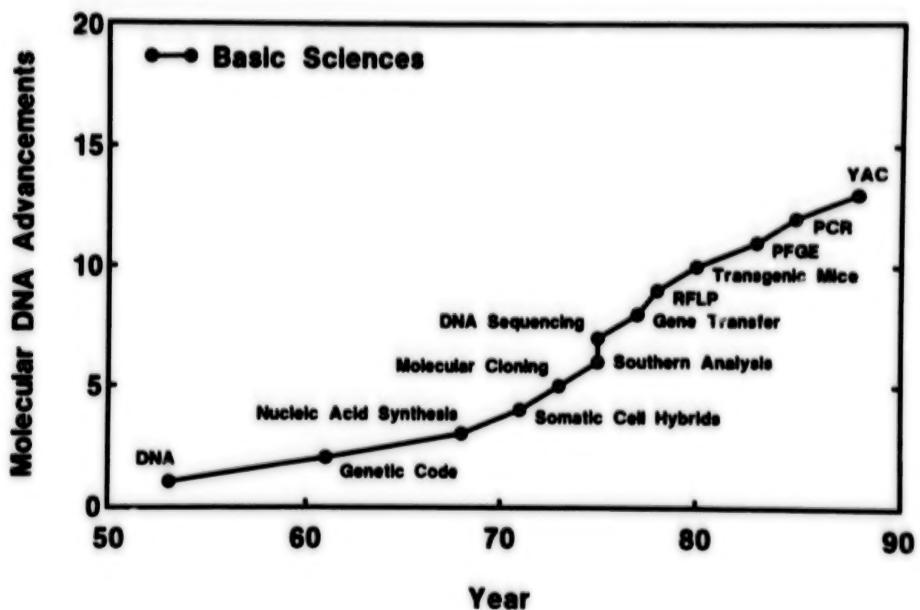
Figure 1. Molecular DNA advancements by year.

tory of DNA diagnostic applications. Forensic scientists are just beginning to use technologies that were developed rapidly and escalated in the late seventies and early eighties. The rapid growth in application of DNA diagnostic in medical sciences can be expected for the forensic field. There is a great deal to be learned and improved in both areas.

DNA VARIABILITY

The variability of DNA is extensive. The base pair (bp) composition of man is quite large at $3 \times$ 10^9 bp per haploid genome. Very little of the human genome is restricted from variation, probably less than one percent since that is the approximate fraction of the genome composed of active, expressed genes. Evidence from our laboratory suggests that even this fraction of the genome is not immune from variation since we have observed the presence of highly polymorphic short tandem repeats (we discuss this in the section on Technologies of the Future) and single base polymorphisms within the exons of expressed genes. So the potential for wide variation from individual to individual
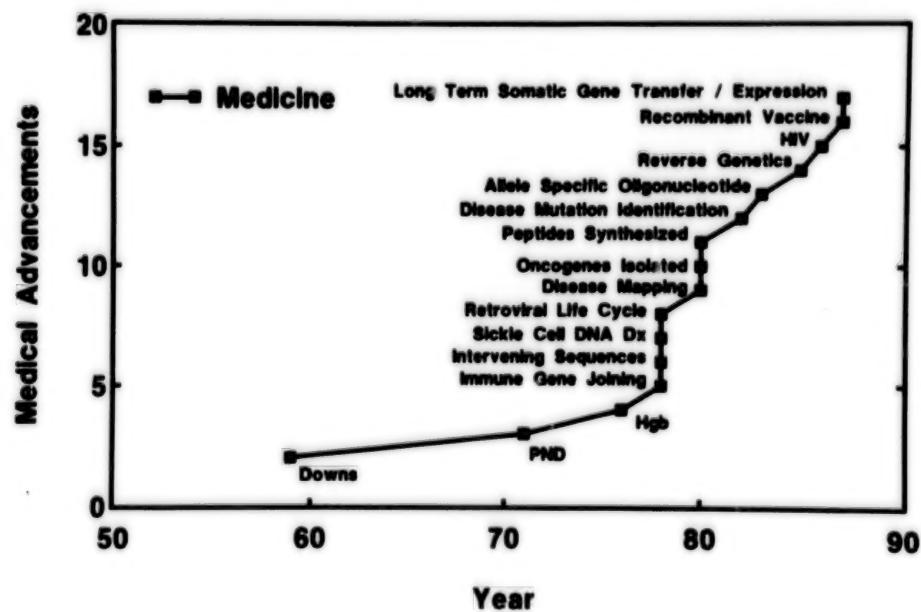
Figure 2. Medical advancements brought about by molecular DNA technologies.

exists and most workers in the field feel that the ability to discriminate between DNA samples derived from different individuals is immense relative to the population size. Regions of the genome between the coding segments of genes (introns) and intergenic segments exhibit DNA sequence polymorphisms ranging from approximately 1/100 bp to 1/1000 bp.

LABORATORY STANDARDS FOR DNA PROFILING

How is this background of technology and the high variability of DNA between individuals translated into the ability to give reliable DNA test results in the court system? Important features of the delivery of this service will include the following: First, the scientific method being employed must be accepted and agreed upon by the scientific community. This feature is not restricted to Southern technology, which is the workhorse of the forensic field at the present time, but also applies to new PCR and sequencing technologies which may prove more powerful and reliable. The same scientific reliability issues must be faced regardless of the method. Second, the laboratory procedures must fulfill acceptable criteria for the field of forensic science—not necessarily those of a molecular biology laboratory where responsibility for disease diagnosis and interpretation of forensic evidence does not occur. The laboratory procedures must be developed by the forensic scientist utilizing the technical expertise of DNA experts from academia and industry. Only the forensic scientist has the practical grasp of situations that occur in forensic casework and understands the limits of applying methods developed for the ideal laboratory setting to the forensic setting. Third, there must be an external quality control system monitoring forensic laboratories. An example is the American College of Pathology (ACP), which serves as an external review panel for medical diagnostic laboratories. The ACP makes unannounced, on site checks on laboratories for laboratory quality control. Furthermore, the ACP supervises the submission of unknown, coded specimens for laboratory analysis and reviews the accuracy of test results. The review is carried out by an external ACP panel. The laboratory reviews are thorough. A laboratory director that experiences an ACP examination typically receives a deficiency report. Those deficiencies are generally small, technical and readily corrected. They, on occasion, identify serious procedural errors. Such deficiencies generally do not discredit the laboratory, but help to maintain high quality control. The American Society of Crime Laboratory Directors might be involved in this external quality control system. Fourth, there are needs to qualify laboratory directors. These laboratory director qualifications might be set by the American Academy for Forensic Sciences. Qualifications for a laboratory director might include examination for DNA intellectual knowledge, laboratory training and supervised experience. Successful passing of such an examination could lead to certification to direct a DNA forensic laboratory.
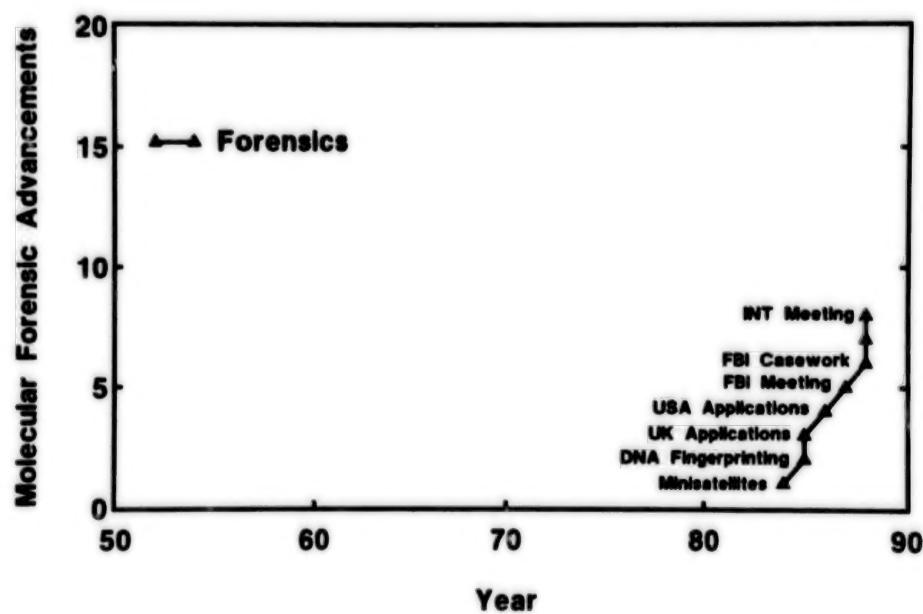
Figure 3. Forensic advancements involving DNA profiling methods.

The methods used to analyze forensic samples must incorporate appropriate controls and standards. For example, there must be adequate population data bases that can be used by individual or national laboratories. This permits an accurate numerical estimate of DNA profiling methods. This data source is critical to the significance of molecular DNA matches. Data bases will be a particular point of discussion in the U. S. given our population mixing. Best estimates will undoubtedly be the standard since absolute precision will be difficult to defend. At the technical level, Southern analyses of suspect(s) and forensic material on the same autoradiogram are ideal. Investigative and forensic case work does not, however, take place as an ideal situation. Some forensic samples may be acquired within a time period which does allow their study on single gels. In other cases, forensic samples may be gathered over wide time periods requiring that analyses be carried out at separate times and on separate gels. In this case, external molecular weight standards and internal sample monomorphic sequences should enable the matching of RFLPs and the detection of aberrations in the migration of DNA fragments gel to gel. Internal controls for the one could employ a set of probes for non-variable human sequences of established molecular weights as sample internal DNA molecular weight references. This internal control would then enable the detection of band pattern distortion occurring in the Southern analysis. There must be defined match/mismatch rules for DNA fragment sizes that is reliably reproduced. Highly informative VNTR probes provide a wide heterogeneity of allele sizes. The assignment of a band to a specific allele is determined by the reproducible resolving power of the agarose gel. The identification of an allele by Southern technology must take into account these reliabilities. This reliability must be clearly defined since it is used in the mathematical estimates of the significance of band matches by referring to population data bases. Because of the likely informative character of VNTRs one should utilize conservative numerical estimates of match for a given probe since the multiplicity of matches with non-linked probes provides the most powerful evidence of match/mismatch of forensic specimens to suspect.

TECHNOLOGIES OF THE FUTURE

The polymerase chain reaction (PCR) is a method that has had a major impact in the area of molecular biology, medical diagnostics and undoubtedly will have a major impact in the field of forensics. The method enables the precise and specific in vitro amplification of DNA fragments ranging in size from approximately 50 bp to 4000 bp in 1 to 5 hrs. (Saiki et al. 1988). The PCR can be applied to DNA profiling in a variety of modes, including DNA sequencing, mismatch detection by chemical cleavage, allele specific probes and amplification of polymorphic units or sites. The DNA sequence of fragments amplified with the PCR can be sequenced directly (Gibbs et al. 1989a and reference therein). Direct DNA sequencing of PCR products enable the accurate determination of the DNA sequence, because rare sequence errors introduced by the thermostable DNA polymerase during amplification are not detected (Gibbs et al. 1989a). The method is most sensitive, rapid and adaptable to automation. Sequencing from PCR products can be performed under standard radioactive sequencing conditions or with the assistance of one of the commercially available fluorescent DNA sequencers (Gibbs et al. 1989a). Figure 4 illustrates the use of direct DNA sequencing of PCR products to detect differences between normal and mutant genes in the clinical and medical diagnostic setting (Gibbs et al. 1989a). Thus, the method can be used to detect differences in the sequence of DNA between individuals—the fundamental basis of all DNA profiling methods.

The PCR can also be used in conjunction with chemical or mismatch cleavage. Mismatch cleavage enables the detection of all classes of DNA mismatches in a heteroduplex molecule (Cotton and Campbell 1989). After formation of a heteroduplex between a reference DNA fragment, (for example, normal) and the test fragment one of the two DNA strands in the heteroduplex is cleaved at the position of any mismatches present in the heteroduplex. One of the two bases in the mismatch is chemically modified, then the DNA strand is cleaved using Maxam-Gilbert chemistry for strand cleavage. Thus, if sequence variation exists in the region, one of the strands will be cleaved and it will be detected by variation in fragment size after electrophoresis on a denaturing polyacrylamide gel. Mismatch cleavage can be greatly simplified with the use of the PCR by amplifying both the reference and test fragment in the test tube to high quantity. The method is being widely applied now in medical diagnostics, linkage and may have a role in the ability to identify individuals. For example, we have used as the method in the Duchenne Muscular Dystrophy region. In a particular 275 bp region six positions show variation and can be used as a haplotyping system based on single based alterations (Gibbs et al. 1989b). Thus, polymorphic sites

6
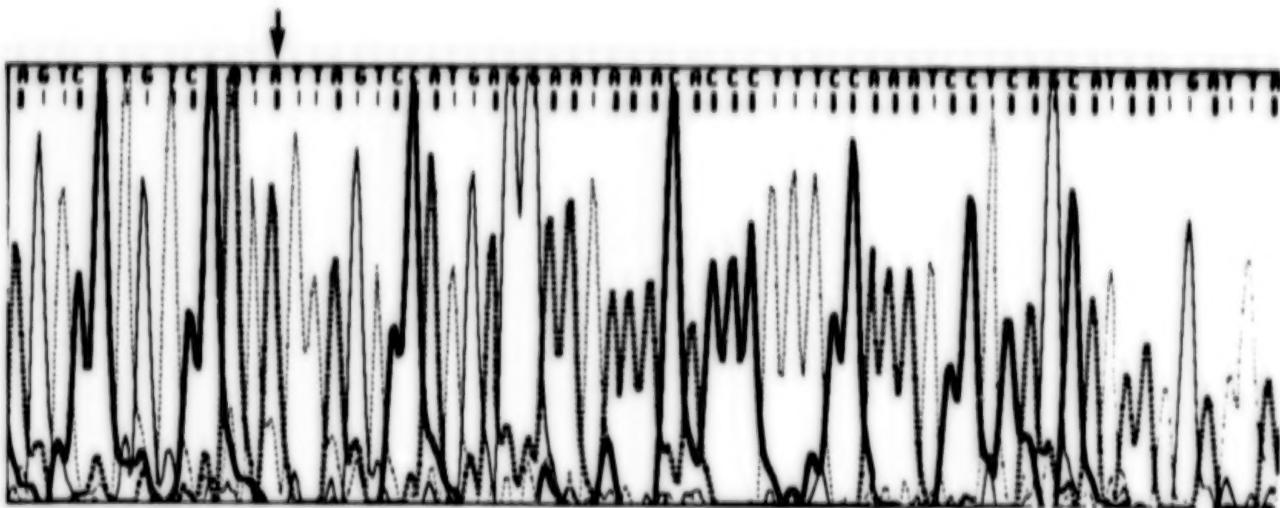
Figure 4. Direct fluorescent DNA sequencing of amplified product from a polymerase chain reaction. The arrow denotes a mutation within the coding region of the human HPRT gene in a patient with Lesch-Nyhan syndrome identified by comparison to the known wild type sequence. The sequence is shown across the top (A, C, G, T).

can be identified by scanning molecular regions for mismatches and subsequently routinely detected by a variety of methods, including molecular weight separation by electrophoresis, allele specific oligonucleotide hybridization and DNA sequencing.

We have recently developed a new method for DNA profiling in our laboratory which is based upon the identification of a novel class of minisatellite sequences or short tandem repeats (STRs) within the human genome. These STRs are a variation of the minisatellite sequences originally reported by Alec Jeffreys (et al. 1985). We have studied both trimeric and tetrameric tandemly repetitive sequences and found them to be highly polymorphic. Figure 5 illustrates one of the tetrameric STRs we have studied. The sequence is (CATT)n where the reiteration number, n, was found to vary from 6 to 10 in a population of greater than 80 individuals. Because the STRs are small they can be amplified faithfully using the PCR and discriminated by electrophoresis on a denaturing gel (DNA sequence gel). This method enables the precise and accurate determination of the alleles. The nucleotide sequences of each of 7 alleles from an STR found within the human HPRT gene [(CTAT)n] has been determined and used to demonstrate that each of the 7 alleles vary only in the reiteration number. Figure 5 also illustrates the use of this DNA profiling method to determine paternity.

Profiling of DNA with STRs will require a number of unlinked allele systems, such as the two just discussed. Based on our experience with four polymorphic STR loci, we estimate that 10 to 15
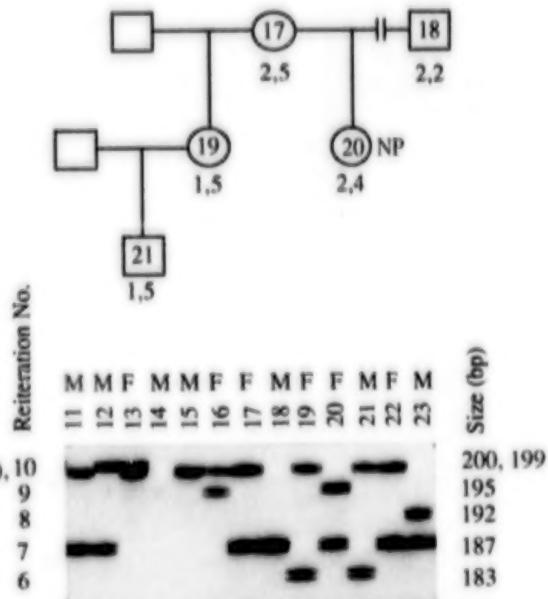
Figure 5. DNA profiling with short tandem repeats (STRs). The data demonstrate that male 18 could not have been the father of child 20. Methods: A tetrameric, tandemly reiterated sequence [(CATT)n] located on chromosome 11 was amplified using the polymerase chain reaction (PCR) as described by the manufacturer (Cetus). The PCR products were labeled with ^{32}P-a-dCTP by adding 2.5 1Ci (3000 Ci/mmol) to the 15 11 PCR. The amplified products were then denatured and fractionated by electrophoresis on a standard sequencing gel. Doublets are seen because both DNA strands are labeled and they have different mobilities during electrophoresis. The reiteration numbers 6 through 10(?) correspond to allele numbers 1 through 6 in the pedigree. The numbers inside the circles and squares of pedigree refer to the lanes in the gel below. Squares represent males and circles represent females.

STR loci would enable one to distinguish between two individuals with a probability of greater than

7

one in one hundred million. How many STR loci are present within the human genome? Examination of human DNA sequences in GenBank suggests that the human genome contains tens of thousands of STRs (Tautz *et al.* 1986; our unpublished results). Thus, the potential for distinguishing between two DNA samples by this method will not be limited by polymorphic loci.

How might DNA profiling with STRs be applied to the forensic sample? One method which will greatly facilitate both the technical execution of the method and interpretation of the results is the multiplex PCR method developed in our laboratory (Chamberlain *et al.* 1988). In this method, multiple fragments are amplified simultaneously with the PCR in a single tube. We have amplified up to nine DNA fragments with the multiplex PCR method. Thus, multiplex PCR provides substantial savings in time and reagents and enables multiple analyses to be performed on the same amount of sample as would be required for one analysis. The multiplex PCR method has been applied to our DNA profiling method utilizing STRs. Multiplexing simplifies the interpretation of the data, since one internal standard suffices for all of the analyses. The small size of the STR amplification units enable the use of the exquisite precision of DNA sequencing polyacrylamide gels for allele determination. Although each of the alleles could be verified by direct DNA sequencing of the amplified products, this should not be necessary on a routine basis once the allele system has been fully characterized. A fluorescent DNA sequencer could be used to advantage, since the multiplexing and incorporation of internal size standards would be simplified. The features of the system (small fragment size and the use of the PCR) also enable the use of degraded and/or limited samples, that could not be used for conventional Southern analysis (Chamberlain *et al.* 1988).

CONCLUSIONS

The extensive variability of DNA within the human genome coupled with the nucleic acid techniques of the molecular biologist are providing the forensic scientist with powerful methods for DNA profiling, whose full potential has not been reached. Despite the criticisms of the current system of DNA analysis system, you have an analytical system which will be upheld on scientific principles and serve the forensic science community and public reliably. Profiling of DNA will benefit the legal system—both defense and prosecution. Operating a high quality laboratory will assure the proper place of DNA technology in the forensic science field. Casual operation of a laboratory will deny this very valuable technology to forensic science. This is a time for forensic scientists to provide the highest quality services for the courts. There is little doubt the technology will continue to improve based on well founded scientific principles.

REFERENCES

Caskey, C. T. and White, R. L. (eds.). (1983). Banbury Report 14: Recombinant DNA Applications to Human Disease. Cold Spring Harbor, NY, Cold Spring Harbor Press, NY.

Chamberlain, J. S., Gibbs, R. A., Ranier, J. E., Nguyen, P-N. and Caskey, C. T. (1988). Detection screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification, Nucleic Acids Res. 16:11141–11156.

Cotton, R. G. H. and Campbell, R. D. (1989). Chemical reactivity of matched cytosine and thymine bases near mismatched and unmatched bases in a heteroduplex between DNA strands with multiple differences, Nucleic Acids Res. 17:4223–4233.

Gibbs, R. A., Nguyen, P-N., McBride, L. J., Koepf, S. M. and Caskey, C. T. (1989a). Identification of mutations leading to the Lesch-Nyhan syndrome by automated direct DNA sequencing of in vitro amplified cDNA, Proc. Natl. Acad. Sci. USA 86:1919–1923.

Gibbs, R. A., Chamberlain, J. S. and Caskey, C. T. (1989b). In: PCR Technology: Principles and Applications for DNA Amplification. Erlich, H. A. (ed.), Stockton Press, NY.

Gill, P., Jeffreys, A. J. and Werrett, D. J. (1985). Forensic application of DNA "fingerprints", Nature 318:577–579.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985). Individual-specific "fingerprints" of human DNA, Nature 316:76–79.

Lander, E. S. (1989). DNA fingerprinting on trial, Nature 339:501–505.

Orkin, S. H. (1986). Reverse genetics and human disease, Cell 47:845–850.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, H., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, Science 239:487–491.

Tautz, D., Martin, T. and Dover, G. A. (1986). Cryptic simplicity in DNA is a major source of genetic variation, Nature 322:652–656.

BLANK PAGE

DNA CHEMISTRY AND GENOME ORGANIZATION: AN INTRODUCTION FOR THE FORENSIC SCIENTIST

Roger Kahn

Metro-Dade Police Department
Miami, Florida

All biological techniques used to identify individuals take advantage of genetic polymorphism. A method based on a genetic system that is highly polymorphic, that is, one for which there are many possible and common alleles, is the most desirable since it best discriminates between individuals. On the other hand, one that reveals very few allelic variants is of little use. With the ultimate individualizing test, no two individuals would share the same result. Identity testing of DNA often produces results that come close to providing unequivocal identification.

To most effectively utilize the powerful technique of DNA identity testing, it is essential to have a good understanding of the molecular structure of DNA, its characteristics both inside and outside of the cell, and its organization in the genome. Fortunately, the structure of DNA and its arrangement in the cell are not difficult to describe or to understand; in fact, the molecule is remarkably beautiful and elegant in its simplicity.

THE MACROMOLECULES

Deoxyribonucleic acid or DNA, together with the nearly identical molecule termed ribonucleic acid (RNA), is one of three classes of macromolecules found in all cells. Like the others, polysaccharides and proteins, the nucleic acids consist of a polymer of repeating subunits. In the case of the polysaccharides, various sugars form the basic units. For proteins, at least 20 different amino acids compose the primary units. In DNA, four types of nucleotides serve as the fundamental subunits.

Although macromolecules all consist of a large number, but limited variety, of specific subunits, they possess very different functions. In animal cells, polysaccharides provide primarily a source of stored energy. Proteins perform a variety of functions including catalysis of chemical reactions, regulation of information flow and service as structural elements. However, the function of the nucleic acids is the most important, for these molecules store genetic information in a coded form, thereby making all other structure and function possible.

The term macromolecule is aptly used to describe these classes of molecules: their molecular weights can all exceed 10,000 daltons. But of the three, DNA is by far the biggest single molecule in a human cell. The molecular weight of the average human DNA polymer is on the order of 60 billion daltons. These giant, linear DNA chains are known as chromosomes; each human somatic cell houses 46 chromosomes in 23 pairs. There are about 100–300 million polymerized nucleotide subunits per chromosome for a total of 12 billion nucleotides per diploid human cell. Overall, the DNA from the chromosome set of a single cell is about 1.5 m long, a number that is remarkable since each nucleotide contributes just 3.4 A to the length. Even more striking, the total length of the DNA in all of the 100 trillion cells of the human body is about 93 billion miles, roughly 1000 times the distance of the Earth to the Sun. While the chromosome is the largest unit of organization of the genetic material, the nucleotide is the smallest. It is the fundamental unit of DNA and any discussion of the human genome should start there.

DNA STRUCTURE

The nucleotides are each composed of three entities: a sugar (deoxyribose in the case of DNA or ribose in the case of RNA), a phosphate group and a base (Figure 1). The DNA polymer is composed of a sugar-phosphate backbone, which is always the same, and the attached bases, which alternate among the four possible ones (Figure 2A). The phosphate groups attach to the sugar in a very specific fashion via an ester linkage (C–O bond), with two bonds per phosphate group. This is described as a phosphodiester linkage. Attached to each sugar at the 1' position is one of four bases: adenine, thymine, guanine or cytosine (Figures 1 and 2B).

NUCLEOTIDES

A nucleotide consists of
a nitrogen-containing
base, a 5-carbon sugar,
and 1 or more phosphate
groups.

BASE

PHOSPHATE
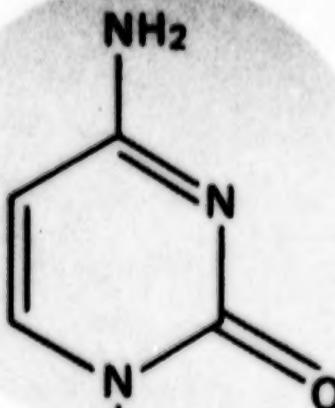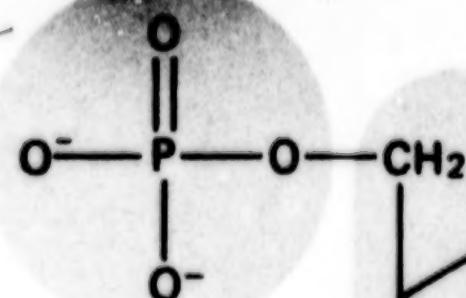
SUGAR

They are the
subunits of
the nucleic acids.

Figure 1. NUCLEOTIDE STRUCTURE. The nucleotide shown is an RNA nucleotide. In DNA, the right-hand hydroxyl group (OH) on the sugar moiety is replaced by a hydrogen atom. Thus, RNA contains ribose while DNA contains deoxyribose. (Alberts *et al.* 1983). Molecular Biology of the Cell. Reprinted with permission from Garland Publishing, Inc., NY.

SUGAR-PHOSPHATE BACKBONE OF DNA
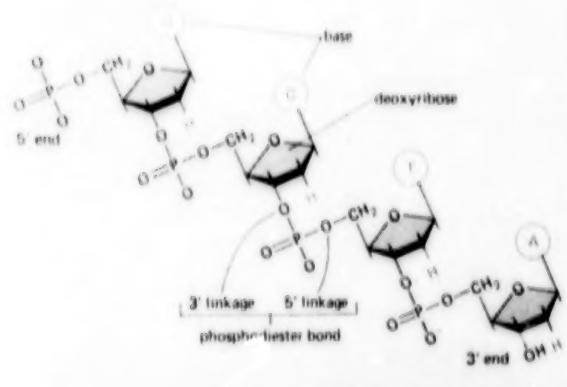
FOUR BASES AS BASE PAIRS OF DNA
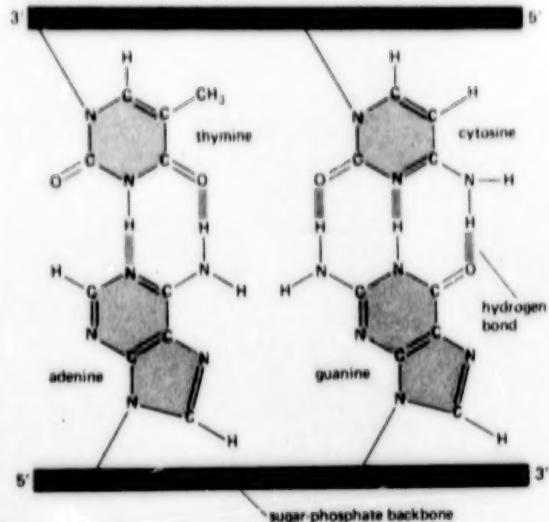
ELECTRON MICROGRAPH OF DNA

DNA DOUBLE HELIX

1 helical turn = 3.4 nm

minor groove   major groove
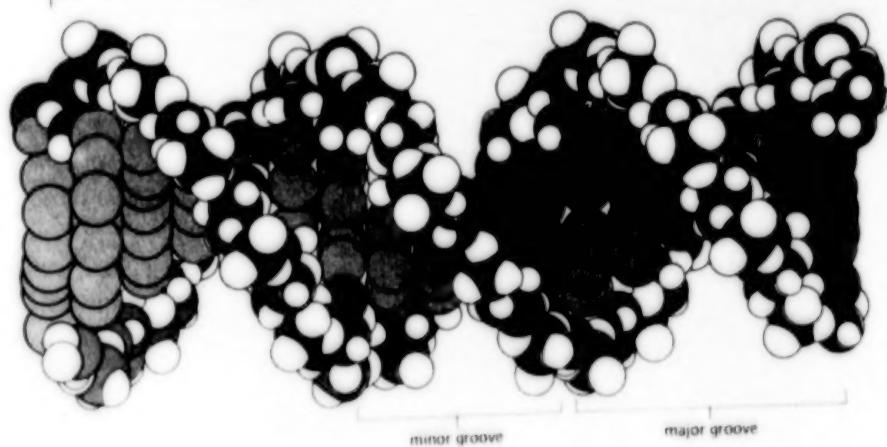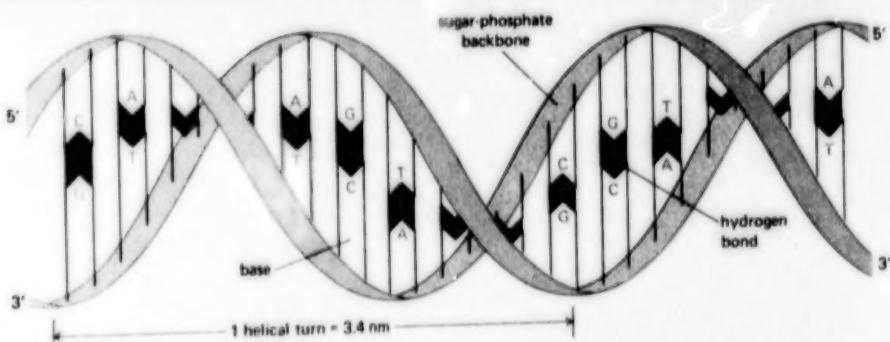
Figure 2. THE STRUCTURE OF DNA. A—Single-stranded DNA polymer demonstrating strand polarity, phosphodiester linkages and the negatively charged phosphate groups of the backbone. B—Structure of the four bases found in DNA as hydrogen bonded base pairs. Note that cytosine-guanine base pairs have three hydrogen bonds while thymine-adenine have 2. C—Schematic diagram of the double-helix: the sugar-phosphate backbone forms the exterior surface while the base pairs face the interior. D—Space-filling model of the double-helix. (Alberts *et al.* 1983). Molecular Biology of the Cell. Reprinted with permission from Garland Publishing, Inc., NY.

13

The bases also bind very specifically to each other: adenine pairs with thymine and guanine pairs with cytosine. This pairing occurs across separate strands of sugar-phosphate backbone, with the result that DNA is generally found in a double-stranded state, one strand wrapped around the other to form a double-helix (Figures 2C and 2D). The two strands, arranged with the backbone on the outside and the bases on the inside, are held together by rather weak hydrogen bonds between bases on opposite strands and by stacking interactions between adjacent bases on the same strand. These stabilizing forces cause the bases to align in a parallel arrangement. Since the pairing of bases is completely specific, the sequence of one strand predicts the sequence of the other.

Each phosphate is linked, on one end, to the 5' carbon of a sugar molecule (the one that is not in the ring) while the other end is attached to the 3' carbon of an adjacent sugar molecule. Thus, the backbone has a polarity: a linear polynucleotide chain that always has both a 5' and a 3' end (Figure 2A). In the double-helix, the two chains run anti-parallel to one another such that the chain running in one direction must pair with a complementary chain running in the other direction (Figure 2C). By convention, DNA sequences are read in the 5' to 3' direction.

DNA PROPERTIES AND HANDLING CHARACTERISTICS

The weak interactions that maintain DNA in the double-stranded state are rather easily disrupted, and the two strands may be unwound and separated. This process, termed denaturation or melting, is an essential part of the most common DNA typing tests in use at this time. High temperature or high pH can be used to promote denaturation, although alkaline conditions are less likely to cause breakage of DNA than is elevated temperature. Low pH will also promote denaturation but must be avoided due to the tendency of DNA to release purine bases at pH 4 or less and to hydrolyze phosphodiester bonds at pH 1 or less. Whereas DNA is stable in alkali to pH 13, exposure to acid can render it useless for analysis.

Denaturation of DNA is reversible. The formation of duplex DNA from complementary chains of single-stranded DNA is known as renaturation, annealing (or reannealing) or hybridization. Strictly speaking, the term hybridization indicates that the reassociated strands were not original partners. The three terms are commonly used interchangeably, however. Unlike denaturation which

can be accomplished in minutes or even seconds, renaturation is a slow process dependent upon the concentrations of the hybridizing sequences. Under conditions typically used for DNA identity testing, it requires several hours.

In addition to extended lengths of time, hybridization requires rather precise conditions of ionic strength and temperature. The temperature must be high enough to disrupt any intrastrand base pairs but low enough to allow the stable formation of interstrand base pairs. The optimal temperature is 20 to 25 °C below the melting temperature.

Increasing the salt concentration stabilizes base pairing and, therefore, increases the tendency of complementary single strands to renature. This effect is due primarily to shielding of the electrostatic charges of the negatively charged phosphate groups by sodium ions and by a decrease in the solubility of unpaired bases at elevated salt levels. Typical salt concentrations in hybridization reactions range from 0.15 M to 1.0 M.

Hybridization reactions are usually performed under conditions which favor formation of duplex DNA. Duplexes that are imperfectly paired can be subsequently eliminated by high temperature and low salt concentration, so-called stringent conditions. Reducing the stringency of hybridization conditions encourages the maintenance of duplexes containing mismatched or otherwise unpaired bases.

Purified DNA can be easily stored in a manner that will prevent its degradation and, in general, DNA in stain materials can be more stable than enzymes or proteins commonly assayed from forensic stains. Purified DNA can be stored for many months, perhaps even years, in a sterile aqueous solution buffered at pH 7 to 9 at 4 °C. It may be stored indefinitely if it is lyophilized or frozen in solution. The DNA in blood or sperm cells that has been allowed to dry into a stain has been shown to last for months if kept in a cool, dry environment at room temperature (R. Kahn, unpublished results; D. Adams, personal communication).

The longevity of DNA in various forensic samples is the subject of on-going investigations in several laboratories (results appear in this volume). In general, however, degradation of DNA in stain materials will be promoted by exposure to UV light or other forms of radiation, to extremes of pH or to wet and humid conditions. The effect of humidity on the degradation of DNA in stains can be at least partly explained by the susceptibility of DNA to attack by a variety of enzymes, known as

nucleases, that are found in all cells. Nucleases are released by microorganisms into their surroundings as a part of their food accumulation effort and by lysing eukaryotic cells. Storage conditions which support the growth of microorganisms will tend to reduce the time that intact, highly polymerized DNA can be recovered from stains. Microorganisms will also contribute DNA to preparations of stain DNA increasing the difficulty of recovering pure samples.

GENOME ORGANIZATION AND INFORMATION FLOW

The Central Dogma of biology states that information flow in a cell follows the path:

transcription translation
DNA ------> RNA ------> PROTEIN

and except for certain interesting, but unusual, exceptions this scheme is ubiquitous (Figure 3). (The diagram does not show the third path of information flow, that of DNA to DNA, which occurs during DNA replication.)

To start the flow of information, DNA is copied or transcribed into the related, single-stranded nucleic acid, RNA, which is then processed in the nucleus. Processing includes (1) the addition of a single, uniquely modified guanine nucleotide cap to the 5' end of the transcript, (2) the addition of approximately 200 adenine nucleotides to the 3' end and (3) the removal of non-coding stretches of RNA which interrupt the coding segments. In eukaryotes, the first process is mandatory, and the second occurs with the vast majority of transcripts. The third process occurs whenever non-coding information is encoded within the gene, as in most genes of higher eukaryotes. The RNA is then transported across the nuclear membrane into the cytoplasm where, on structures known as ribosomes, the nucleotide code is translated into the corresponding amino acid sequence (Figure 3).

There are, in fact, four classes of RNA sequences involved in the sequential phases of information flow. One of these, messenger RNA or mRNA, contains the code that will be translated precisely into the amino acid sequence of a protein. The three other classes, namely, ribosomal RNA (rRNA), transfer RNA (tRNA) and small nuclear and cytoplasmic RNA (snRNA and scRNA), serve specific functions as RNA molecules and are not translated into protein.

There are several species of rRNAs and they are integral parts of the ribosome, serving a structural as well as a mRNA stabilizing function during the initiation of translation. The tRNA molecules are adapters which convert nucleic acid sequences into amino acid sequences according to the triplet codon rules of the genetic code. They accomplish this by pairing in specific, complementary fashion to the mRNA in the ribosome while carrying a specific amino acid to add to the protein under construction. There are usually veral tRNA molecules for each of the twenty amino acids due to the redundancy of the genetic code. Both snRNA and scRNA compose a class of molecules whose functions are only partly elucidated. Some play a role in the excision of non-coding sequences of mRNA and in mRNA transport into the cytoplasm. They may have additional roles in stabilizing mRNA and other as yet undeciphered functions.

As mentioned previously, most of the genes of higher eukaryotes, including man, contain internal sequences that are spliced out of primary transcripts before they are transported out of the nucleus. The excised segments are termed introns or intervening sequences and the coding sequences, which are spliced together to form the functional mRNA, are called exons. Introns are found in rRNA and some tRNAs as well as in mRNA.

Introns in mRNA, however, frequently comprise the majority of the length of the original transcript, and in some cases comprise nearly all of the primary transcript. An extreme example of this seen is found in the gene for human Factor VIII which is involved in blood clotting. This gene is approximately 186,000 base pairs in length, but after the removal of 25 introns, the mature transcript consists of only about 9000 nucleotides (Gitschier et al. 1984). The exons vary in size from 69 to 3106 base pairs; however, all but two are less than 320 base pairs. This motif of relatively small exons bounded by remarkably large introns is not unusual. However, the number of introns and the number of nucleotides per intron vary widely. Some genes, such as histones, contain no introns. Although the actual proportion of intron to exon DNA in the eukaryotic genome is not known, the emerging view is one of large blocks of unique-sequence DNA that do not code for protein. These blocks are found not only as introns, but also as spacer regions of unique sequence DNA lying between genes.

Like intron DNA, spacer DNA can markedly increase the amount of DNA in the eukaryotic genome beyond the minimally required sequences. One of the clearest examples of this phenomenon is the beta-like globin gene family on human chromosome 11. This gene cluster of similar but non-iden-

15
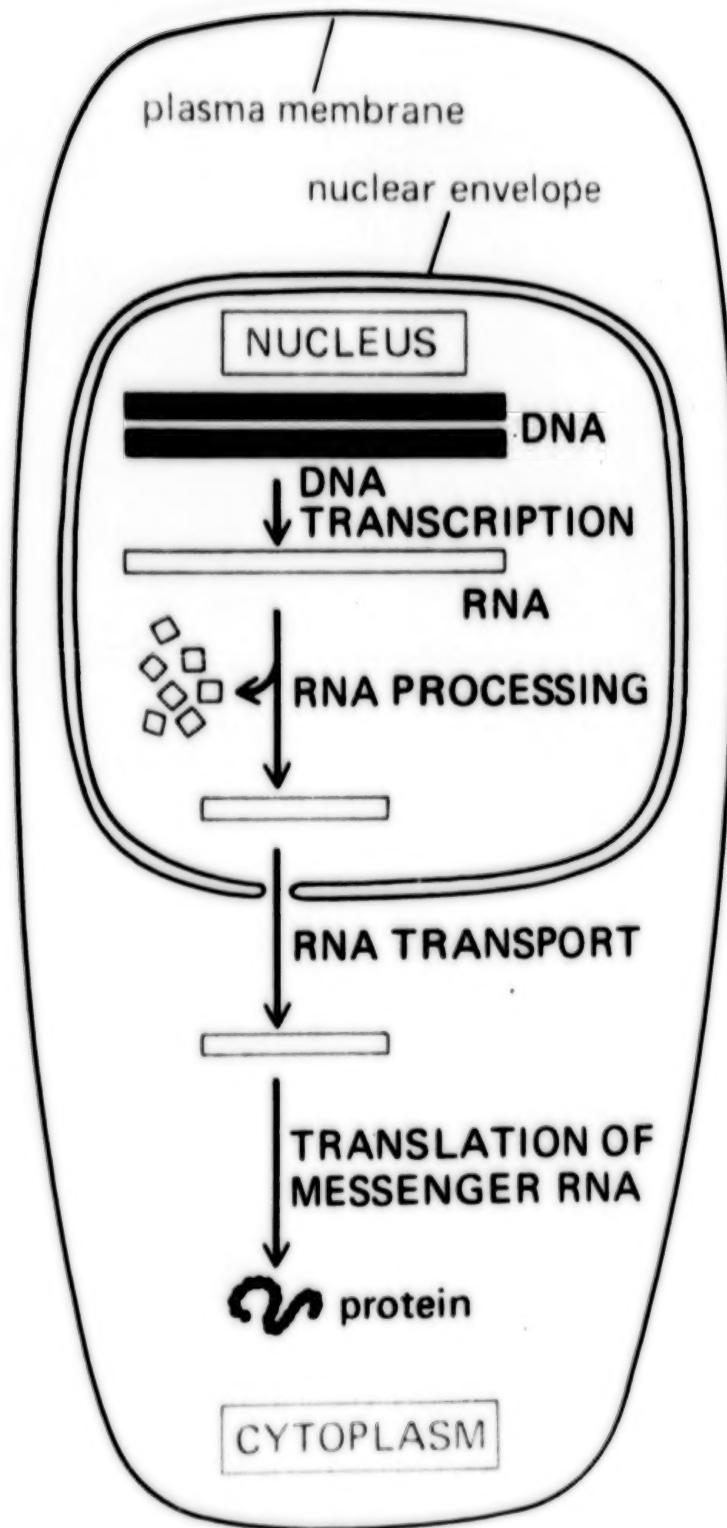
Figure 3. INFORMATION TRANSFER IN THE EUKARYOTIC CELL. Information is transferred from DNA to RNA to protein in a set of discrete, compartmentalized steps. (Alberts *et al.* 1983). Molecular Biology of the Cell. Reprinted with permission from Garland Publishing, Inc., NY.

tical genes spans a region of over 50,000 base pairs. It consists of a tandem array of five, functional globin genes and one dysfunctional globin gene-remnant called a pseudogene. Of the 50 Kb pairs, less than 20% is comprised of active transcription units and, of this, only about 30% consists of sequences ultimately translated into protein. Overall, less than 5% of the DNA in the cluster is translated. A similar motif is found for the alpha-like globin family on chromosome 16, where a 28 Kb pair region contains three active genes and three pseudogenes.

Gene families are not unique to the globin system. Clustering of related genes is especially common for structural proteins such as the tubulins and actins and where marked diversity is desirable such as the variable chain immunoglobulin genes (V genes) and the transplantation or histocompatibility genes (HLA genes). In addition, the genome also contains tandemly repeated genes. Unlike a family, these are a set of identical genes and spacer elements arrayed in a head-to-tail fashion. The genes for tRNAs, rRNAs and the histones are all tandemly repeated, hundreds of times in some organisms.

UNIQUE SEQUENCE VERSUS REPETITIVE SEQUENCE DNA

Hidden in the approximately 3 billion base pairs of DNA of the haploid human genome is an estimated 25,000 to 100,000 genes with an approximate average length of 5 to 10 Kb pairs, including introns. In addition, DNA contains signal sequences for a variety of processes such as gene expression regulation, DNA replication, chromosome packing and chromosome segregation. Including spacer DNA, signal sequences and protein coding

DNA, unique sequences account for approximately 70 to 80% of the total human genome. The remaining 20 to 30% of the genome consists of repetitive, primarily non-coding sequences of uncertain function.

Repetitive DNA can be divided into two classes: tandemly repetitive sequences and interspersed repeats (Table 1). For a recent review and bibliography see Fowler et al. 1988. Tandem repetitive DNA accounts for roughly one-third of the repetitive DNA in the human genome, about 10% of the entire length. Like tandemly repeated genes, these regions consist of head-to-tail repeats of a particular sequence; but they consist of much shorter units, ranging from 5 to 250 base pairs, repeated typically hundreds of thousands of times. Regions of this type are referred to as satellite DNA because of the tendency of some of these sequences to form satellite bands in equilibrium density gradients, thus simplifying their separation from other chromosomal DNA. In humans, as in a variety of higher eukaryotes, there are several classes or families of satellite DNAs; at least some of the classes are associated with the centromeric regions of the chromosomes.

A minor fraction of tandemly repeated DNA, consisting of regions much shorter than satellite DNA, exhibits an extraordinary variability in overall length (Table 1) (Wyman and White 1980). These regions, termed minisatellites (Jeffreys et al. 1985) or variable number tandem repeats (VNTRs) (Nakamura et al. 1987) are comprised of short, 9 to 64 base pairs core sequences in tandem arrays at unique chromosomal loci (Nakamura et al. 1987; Wong et al. 1987). The length of the region depends upon the number of times the core sequence is repeated.

Table 1. REPETITIVE DNA IN THE HUMAN GENOME

REPETITIVE DNA (20 to 30% of the human genome)			
Tandemly repeated sequences (~10% of genome)		Interspersed elements (~15 to 20% of genome)	
Satellite DNA	Short tandem repeats for example, "minisatllites" or VNTRs	SINES (<500 bp) for example, Alu (3 to 6%)	LINES (>500 bp) for example, L1 (1 to 2%)
Satellite DNA sequenes found typically at, or near, chromosome centromeres; regions usually millions of base pairs in length. VNTR regions typically hundreds to thousands of base pairs in length.		Single units scattered throughout the genome—likely to be mobile genetic elements.	

(Modified from Fowler et al. 1988). Copyright ASTM, Philadelphia, PA.

It is possible to determine the lengths of the DNA comprising the two alleles carried by an individual by using standard molecular biological techniques. Testing of this type has revealed re-

markable heterozygosities, greater than 70% at many loci and some as high as 99% (Nakamura *et al.* 1987; Wong *et al.* 1987). These loci have received a great deal of attention from forensic scientists recently as they form the basis of the most common form of DNA identity testing currently in use.

The largest class of repetitive DNA consists, not of tandem arrays, but of individual units of specific sequences distributed widely throughout the eukaryotic genome. In humans, these interspersed elements account for as much as 20% of the total DNA (Table 1). The most common element of this type is termed the Alu family (named for the restriction enzyme Alu I since these sequences generally contain a site for cleavage by this enzyme). These sequences, which are about 300 base pairs in length, account for roughly 5% of the human genome. In other words, a member of this family will be encountered, on the average, every 5 to 10 Kb pairs. Alu DNA and other interspersed repeats that are less than 500 base pairs are categorized as short interspersed elements or SINES. Long interspersed elements, or LINES, are greater than 500 base pairs and are exemplified by the L1 or Kpn family (for the restriction enzyme KpnI) which comprises 1 to 2% of the genome. The L1 elements are found as 7 Kb pair elements or truncations of the parent sequence family. Interspersed elements, as a class, appear to propagate themselves throughout the genome as mobile genetic elements or transposons.

SUMMARY

The advent of DNA identity testing has resulted in a surge of interest in the forensic community in the properties of DNA and its organization in the human genome. This article is intended to provide a framework from which the forensic scientist can pursue these topics in more detail. For those who desire an expanded presentation, a variety of recent texts are available including Alberts *et al.* 1983, Darnell *et al.* 1986, Freifelder 1987, Lewin 1987, Singer 1985, Suzuki *et al.* 1986, Watson *et al.* 1987.

ACKNOWLEDGMENTS

BIBLIOGRAPHY

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. D. (1983). Molecular Biology of the Cell. Garland Publishing, Inc., NY.

Darnell, J., Lodish, H. and Baltimore, D. (1986). Molecular Cell Biology. Scientific American Books, Inc., NY.

Fowler, J. C. S., Burgoyne, L. A., Scott, A. C. and Harding, H. W. J. (1988). Repetitive deoxyribonucleic acid (DNA) and human genome variation—a concise review relevant to forensic biology, J. Forensic Sci. 33:1111–1126.

Freifelder, D. (1987). Molecular Biology. Jones and Bartlett Publishers, Inc., Boston.

Gitschier, J., Wood, W. I., Goralka, T. M., Wion, K. L., Chen, E. Y., Eaton, D. H., Vehar, G. A., Caron, D. J. and Lawn, R. M. (1984). Characterization of the human factor VIII gene, Nature 312:326–330.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985). Hypervariable "minisatellite" regions in human DNA, Nature 314:67–73.

Lewin, B. (1987). Genes. John Wiley and Sons, NY.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987). Variable number tandem repeat (VNTR) markers for human gene mapping, Science 235:1616–1622.

Singer, S. (1985). Human Genetics. W. H. Freeman and Company, NY.

Suzuki, D. T., Griffiths, A. J. F., Miller, J. H. and Lewontin, R. C. (1986). An Introduction to Genetic Analysis. W. H. Freeman and Company, NY.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. and Weiner, A. M. (1987). Molecular Biology of the Gene. Benjamin/Cummings Publishing Company, Inc., Menlo Park, California.

Wong, Z., Wilson, V., Patel, I., Povey, S. and *Jeffreys, A. J. (1987).* Characterization of a panel of highly variable minisatellites cloned from human DNA, Ann. Hum. Genet. 51:269–288.

Wyman, A. R. and White, R. (1980). A highly polymorphic locus in human DNA, Proc. Nat. Acad. Sci., 77:67–54–6758.

DNA MARKERS IN FORENSIC APPLICATIONS

Ray White, Yusuke Nakamura, Kentaro Kasai, Shannon Odelberg and Roger Wolff

University of Utah School of Medicine
Salt Lake City, Utah

The first few years of practical experience with DNA markers in the world of forensic applications have revealed several important features. First and foremost, the promise of the technology had remained solid; nothing has emerged to seriously challenge the idea that molecular probes for highly polymorphic loci in the human genome can reveal a pattern of DNA sequence variation that is sufficiently complex to uniquely identify any individual. However, it is also apparent that the experimental realities of accurate definition of the DNA sequence variants, and the logistics of error-free handling of samples, have not yet progressed to the point that full advantage can be taken of this potential.

The logistical problems that have arisen are familiar in many areas of laboratory operation and are shared by commercial medical testing laboratories and academic laboratories. People make mistakes; samples are switched, gel patterns are misread. A frequency of .0001 might serve as a rule of thumb with respect to human error in DNA technology; this is the frequency of human error in copying DNA sequence data and the frequency of individual pilots landing planes with wheels up.

Two ways of dealing with human error are well understood; automation and redundancy. The DNA sequence-copying problem can be solved by Xeroxing DNA sequences instead of transcribing them by hand. The wheels-up problem is essentially solved by redundancy in the cockpit, two pilots to check and recheck. Laboratory protocols that include astute redundancies and automation of sample handling will certainly help to overcome some of the problems inherent in DNA testing, even though the technology requires a high degree of human participation.

Good DNA protocols run in responsible laboratories should permit very high levels of accuracy for forensic analysis. Furthermore, in cases of doubt, or perhaps even routinely, parallel samples could be run in independent laboratories—perhaps one laboratory chosen by the defense and another by the prosecution. Such redundancy would greatly reduce the chance that an error in sample handling be undetected; the primary challenge of course is to know when an error has occurred.

Our laboratory in Utah has been involved in this effort through development of a series of DNA-based genetic markers. In order to establish chromosomal locations of genes responsible for human genetic diseases, we have developed several hundred DNA markers and arrayed them into linkage maps for each human chromosome (Human Gene Mapping 9, 1987). With the recognition that human individuation was likely to be an important application of the technology used for constructing chromosomal maps, we surveyed the marker systems we had developed to identify those that seemed especially promising for efficient individuation.

We initially reasoned that, just as for application to genetic linkage problems, the ubiquitous restriction-site polymorphisms found in human DNA have a serious liability for human individuation. In principle only two alleles can exist; the site is either present or absent. Individuation, therefore, requires the use of many such systems jointly in order to have sufficient power of discrimination. For example, even in the most favorable case—allele frequencies of 0.5—five systems would have to be developed to yield a discrimination level (a likelihood of matching by chance) of 1/1,000 among unrelated individuals. Such systems, however, do have the advantage that alleles are usually well resolved and unambiguous (Figure 1).

Fortunately, another class of DNA-sequence polymorphism is present in the human genome; a locus of this type carries a tandemly repeated oligonucleotide a sequence, and the number of repeats can be highly variable within the population. We have called these loci VNTRs (for Variable Number Tandem Repeat) (Nakamura *et al.* 1987, 1988); as a matter of interest, a VNTR—whose structure was not recognized at the time—was indeed the first arbitrary DNA-sequence polymorphism identified in the human genome, by Arlene Wyman as a Fellow in my laboratory. Such polymorphic systems escape the two-allele limitation by revealing the different fragment-size classes that reflect the highly variable number of copies of the tandem repeat that can be present at a VNTR locus (Figure 2). Careful electrophoresis of our best system, YNH24, has revealed 67 alleles in a
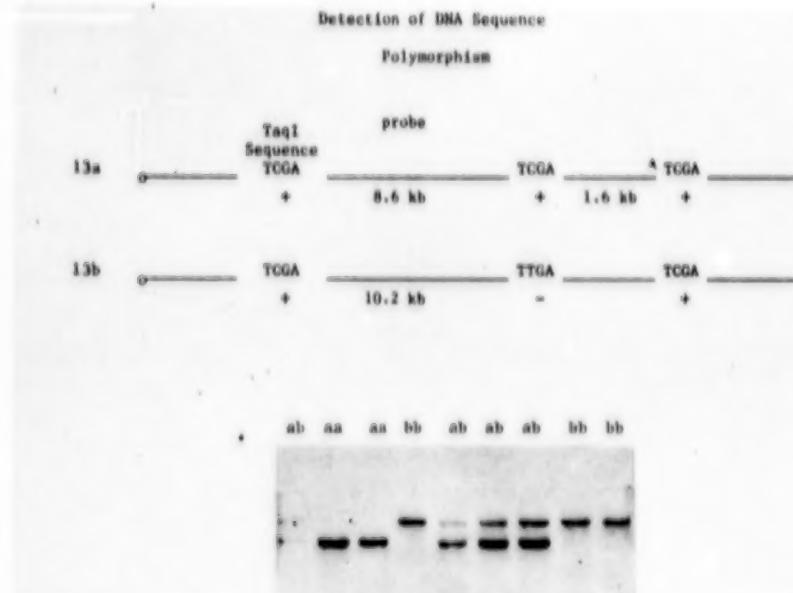
Figure 1. Site polymorphism on chromosome 13, revealed by autoradiography of a Southern blot after TaqI digestion of DNA from nine individuals. The TaqI recognition sequence, TCGA, has been altered by a T>C substitution in allele b at the locus defined by the specific probe, resulting in a longer restriction fragment (10.2 Kb) that is clearly distinguishable from the 8.6 Kb fragment derived from allele a.

population sample; the average probability of a chance match is only 1/1,000 with this system alone.

The inheritance patterns of alleles at VNTR loci have been quite stable in our hands. When gathering genotypic data from families for linkage studies we routinely examine almost 1,000 meioses. Characterizing more than 100 VNTR systems in this sample set, we have detected the emergence of only two new alleles; both of these were at one locus (YNZ22), indicating a frequency of 0.2% for that system and less for the others (Wolff et al. 1988, and unpublished data).

Resolution and interpretation of alleles in VNTR systems, however, presents challenging problems. The resolution achievable by electrophoresis in agarose gels is related to the molecular size of the fragments: the larger the fragments, the larger the size difference needed for resolution. This means that although size differences of 20–50 base pairs (bp) might allow discrimination of fragments in the range of 1000 bp (1 kilobase, Kb), those same size differences might not be detectable among fragments as large as 5 Kb. This is an operationally important range of resolution, because the majority of VNTR repeats are 15–75 bp long. The scarcity of marker systems that provide both good discrimination and good resolution of discrete alleles has created complexities for the interpretation of autoradiographs and the statistics of individuation.

It would be helpful, therefore, if we could assign a unique designation to each allele in a VNTR system. Because the individual alleles of a VNTR system differ by an integral number of repeating units, to fully resolve all possible alleles we must be able to resolve restriction fragments that differ in length by only one repeating unit. We come closest to this goal by creating the smallest possible restriction fragments to define the alleles. In fact, the more resolving power we can achieve, the more robust will be our experimental system, with respect to a decreasing sensitivity to artifacts introduced by conditions within the gel.

A first step in the right direction is to make an intelligent choice of restriction enzymes. In general, the VNTR sequences themselves are free of sites that are recognized for cleavage by restriction enzymes. Enzymes such as HaeIII and HInfI, with four-base recognition sequences that result in frequent cuts in human DNA, have the best chance of serving our purpose. Enzymes that recognize relatively infrequent sites, such as PstI or EcoRI, represent a poor choice because they yield unnecessarily large fragments. However even with an enzyme such as HaeIII, the fact that a locus often contains many copies of the repeating unit means that we will obtain relatively large restriction fragments with most VNTR systems.

How much discrimination can be achieved with the proper choice of enzymes? To provide
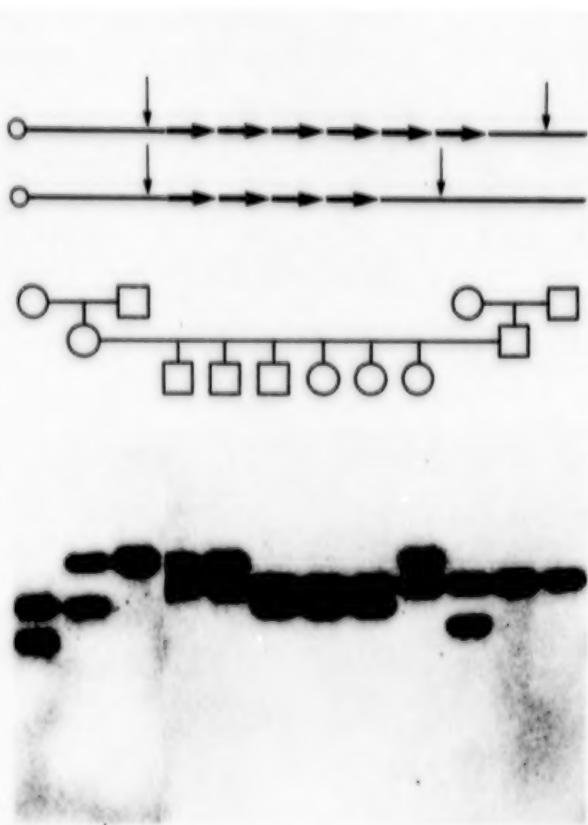
Figure 2. VNTR polymorphism. The autoradiograph reveals four different alleles at the insulin locus in a single family, as a consequence of different numbers of tandem repeats of an oligonucleotide sequence. The insertion/deletion of the repeat elements is shown schematically above the pedigree. (c) 1986, Alan R. Liss, Inc. Reprinted with permission.

some basis for estimation, we have run an initial survey of allele frequencies for eight of our better VNTR systems, obtained with 4-cutter enzymes on a set of 75–151 unrelated individuals from our collection of DNA samples from Utah reference families, testing the power of exclusion for paternity (Odelberg et al. 1989). As Table 1 shows, the average probabilities for matching of random individuals range from 0.001 to 0.09 for the eight systems. The combined average probability of matching two unrelated individuals, therefore, would be of order 10^{-14}. So in principle the power of individuation is enormous, but such numbers must be taken with a grain of salt. For example, in Utah with a population of less than two million and an average sibship size of 4–5, the probability that two randomly drawn individuals are sibs is approximately 1/200,000. The probability of matching two sibs with the full set of eight marker systems is 10^{-4}. Taken together, the likelihood that a random draw will yield two matching individuals is, therefore, close to 10^{-9} if we include the possibility of drawing sibs.

Table 1. MATCHING PROBABILITIES FOR FORSENIC INDIVIDUALIZATION [a]

Locus (Probe)	Enzyme	Probability of Matching	
		2 Random Persons	2 Siblings
D17S5 (pYNZ22)	Hinfl	0.038	0.335
D2S44 (pYNH24)	Hinfl	0.001	0.263
D9S7 (pEFD126.3)	Hinfl	0.035	0.330
D14S13 (pMLJ14)	Hinfl	0.005	0.277
D19S20 (pJCZ3.1)	Hinfl	0.064	0.369
D16S83 (pEKMDA2-I)	AluI	0.030	0.322
D1S74 (cYNA13)	AluI	0.009	0.287
D3S42 (pEFD64.1)	Hinfl	0.092	0.398
All 8 Systems jointly		1.29 x 10^{-14}	1.09 x 10^{-4}

[a] From Odelberg et al. (1989). Used by permission of Academic Press.

In addition to aspects of population structure, familial relationship or problems of allelic resolution can combine to reduce the overall power of discrimination of our marker systems. For example, several of our systems show an excess of apparent homozygotes, to yield significant departures from Hardy-Weinberg equilibrium (Table 2). Whether

this underrepresentation of heterozygotes is due to the underlying population structure or to our inability to resolve the alleles of some heterozygotes is not clear.

Locus (Probe)	Enzyme	Number of Alleles	Number of Persons	Hom.		Het.		X^2	P value
				Obs.	Exp.	Obs.	Exp.		
17S5 (pYNZ22)	*Hin*fl	14	151	21	22.9	130	128.1	0.18	0.649
D2S44 (pYNH24)	*Hin*fl	67	151	8	3.8	143	147.2	4.85	0.014[*]
D9S7 (pEFD126.3)	*Hin*fl	16	136	24	19.4	112	116.6	1.25	0.220
D14S13 (pMLJ14)	*Hin*fl	30	82	12	4.3	70	77.7	14.86	0.0003
D19S20 (pJCZ3.1)	*Hin*fl	13	84	16	17.3	68	66.8	0.11	0.685
D16S83 (pEKMDA2-I)	*Alu*I	15	78	8	10.1	70	67.9	0.49	0.397
D1S74 (cYNA13)	*Alu*I	22	78	10	5.4	68	72.6	4.30	0.021[*]
D3S42 (pEFD64.1)	*Hin*fl	15	84	18	21.0	66	63.0	0.57	0.450

[a] From Odelberg *et al.* (1989). Used by permission of Academic Press.
[*] Significant deviation of observed homozygote and/or heterozygote frequencies from expected frequencies: $P < 0.05$.
Hom. = Homozygotes; Het. = Heterozygotes; Obs. = Observed Frequencies; Exp. = Expected Frequencies.

As already mentioned, we recognize the difficulties associated with the interpretation and statistical analysis of the continuous allele distributions obtained for most VNTR systems; that is, discrete alleles are not always resolved. We have, therefore, reviewed our several hundred VNTR systems to identify those with the smallest size range of alleles, without requiring that they have the largest number of alleles. Ten candidate systems have been identified in which the largest alleles are smaller than 3 Kb but the number of alleles still averages ten, each present at reasonable frequency in the population.

One of these, the YNZ22 locus, has revealed 14 alleles following digestion with *Hin*fl, with an average probability of 0.04 of matching two randomly chosen individuals from our Utah population. The *Hin*fl fragments range in size from 770 to 2800 bp, and the repeat unit is 70 bp. The flanking sequences have been obtained by Roger Wolff in our group and used to develop primers for the polymerase chain reaction (PCR), a procedure now routine for amplifying alleles in semen samples. Upon Southern transfer of *Hin*fl-digested DNA, individual alleles of YNZ22 can be specified in terms of the number of random repeats present. Even in the larger fragments, alleles differing by only one repeat can be resolved (Figure 3). In our set of 151 unrelated, test individuals, almost all alleles observed were smaller than 2 Kb; the largest fragment seen, 2800 bp long, was represented by only one individual (unpublished data).

Alleles with only one repeat unit also appeared in our survey; it is worth pointing out that such alleles would easily be missed due to faintness of the label. In ordinary forensic matching, this might not be critical; but in circumstances where familial relationships might become part of the argument, overlooking such an allele could be very misleading.

In addition to their capacity to provide resolvable, discrete alleles, the new markers offer two advantages. Firstly, because large fragments are the most sensitive to degradation, systems with small alleles are likely to be more informative in forensic samples that have suffered abuse. Secondly, and not unrelated, systems with a small range of allele sizes are the most amenable to amplification by PCR. It has been our experience, and that of others, that fragments more than a few Kb long are likely to be substantially underrepresented, even by orders of magnitude, than fragments of 3 Kb or less. The ten systems we have initially identified therefore become good candidates for PCR as well.

The first we have chosen to develop, MCT118, showed 11 alleles in Southern transfers. The DNA sequencing of MCT118 by Dr. Kasai of the National Research Institute of Police Science in Japan, who has been working in our laboratory with Dr. Nakamura for the past year, revealed 16-base repeats and flanking sequences. A pair of primers, 28 and 29 bp long respectively, were used for 35 rounds of PCR amplification with less than 2 ng of DNA template. An ethidium-stained polyacrylamide gel (Figure 4) revealed unique visible bands, with none of the laddering artifact that is commonly seen after PCR amplification of VNTR systems. The alleles found among a set of 67 unrelated Caucasian and Japanese subjects ranged in length from 384 to 656 bp. In this size range, DNA sequencing gels are able to resolve fragments that differ by only one bp; they are not at all challenged when fragments differ by 16 bp.
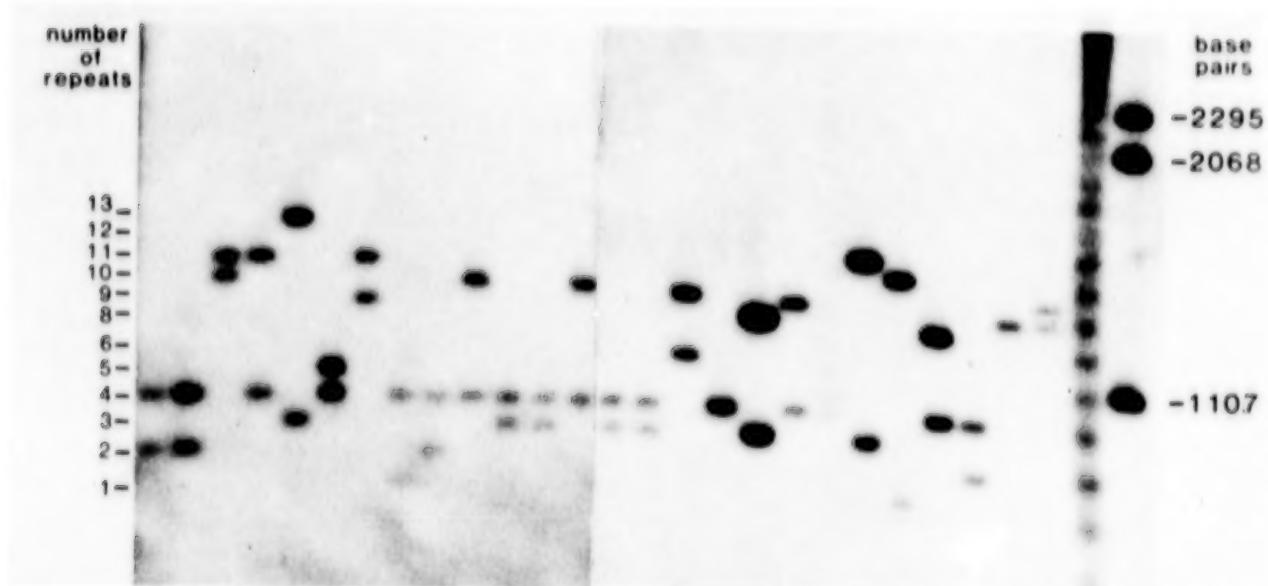
Figure 3. Autoradiograph of alleles at locus YNZ22, obtained from a high-resolution gel. DNA was collected from 25 unrelated individuals, digested with HInfI, and probed with pYNZ22. Twelve of the 14 HInfI alleles are visible in this sample. Far right lane, pBR322 markers; second lane from the right, BRL 123-base-pair ladder. Numbers in the far left column indicate the number of repeating units in each DNA fragment. (Odelberg et al. 1989). Reprinted with permission from Academic Press, NY.
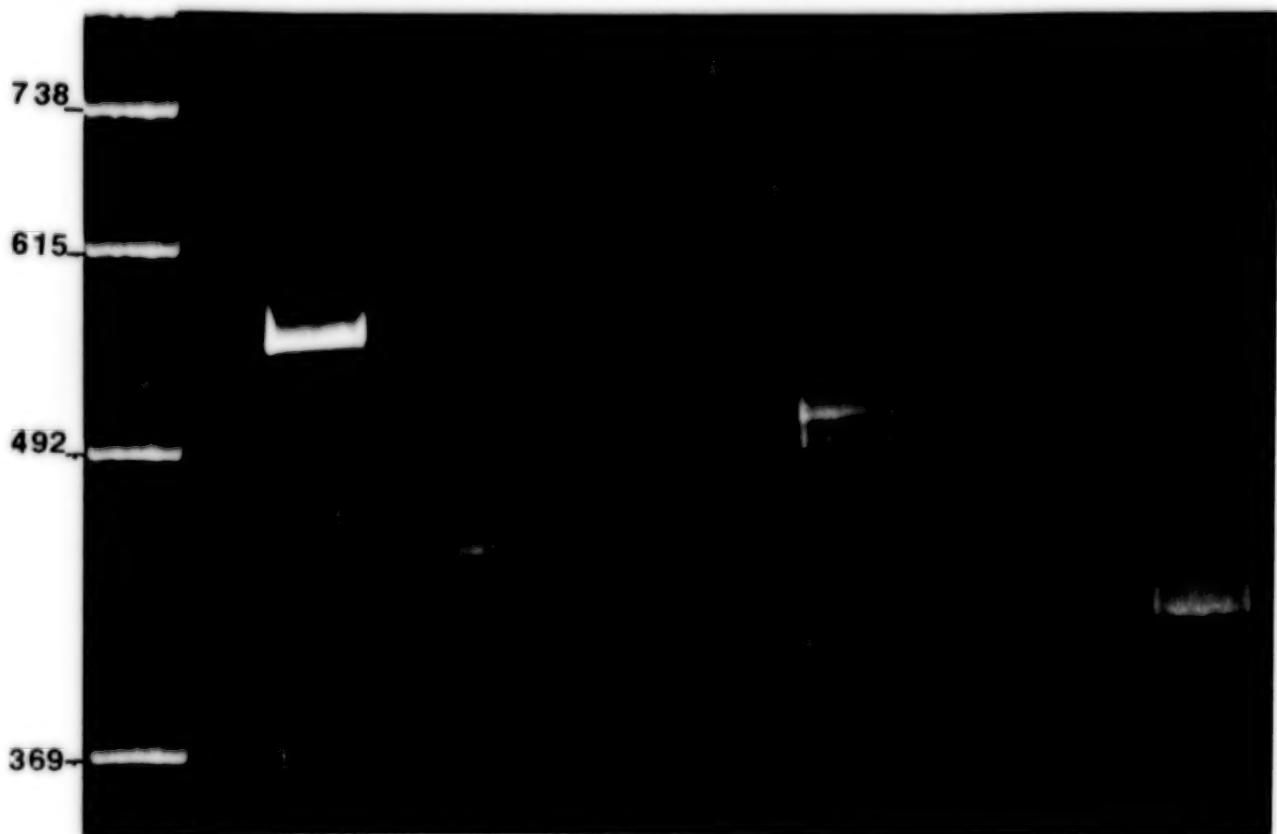
Figure 4. Ethidium bromide-stained polyacrylamide gel following electrophoresis of amplified MCT118 from six individuals. Size markers (far left lane) are a 123 bp ladder.

23

In summary, good progress is being made toward defining new VNTR systems for which individual alleles can be uniquely and robustly specified. Two of these, YNZ22 and MCT118, are now ready for forensic application and can be amplified by PCR if desired. Two more are almost ready. This second generation of systems from our Utah laboratory should reduce many of the complexities of interpretation and statistics that were inherent in the first generation of markers developed by ourselves and others. Similarly, the TGTG systems under development by Jim Weber and Mike Litt offer good prospects for resolution and discrimination. The technical future seems assured; improved marker systems and protocols, coupled with redundancy in testing, should go a long way toward generating a practical level of individuation that approaches the theoretical.

REFERENCES

Human Gene Mapping 9, Paris Conference (1987). Cytogenet. Cell Genet. 46:1–762.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping, Science 235:1616–1622.

Nakamura, Y., Carlson, M., Krapcho, K., Kanamori, M. and White, R. (1988). New approach for isolation of VNTR markers, Amer. J. Hum. Genet. 43:854–859.

Odelberg, S.J., Plaetke, R., Eldridge, J.R., Ballard, L., O'Connell, P., Nakamura, Y., Leppert, M., Lalouel, J.-M. and White, R. (1989). Characterization of eight VNTR loci by agarose gel electrophoresis. Genomics 5:915–924.

Wolff, R.. Nakamura, Y. and White, R. (1988). Molecular characterization of a spontaneously generated new allele at a VNTR locus: no exchange of flanking DNA sequences, Genomics 3:347–351.

RECENT REVIEWS

White, R. and Lalouel, J.-M. (1987). Investigation of genetic linkage in human families, Adv. Hum. Genet. 16:121–228.

White, R. and Lalouel, J.-M. (1988). Sets of linked genetic markers for human chromosomes, Ann. Rev. Genet. 22:259–279.

THE EXTRACTION, PURIFICATION AND QUANTIFICATION OF DNA

F. Samuel Baechtel

FBI Laboratory
Quantico, Virginia

The intent of this document is to give the reader an overview of the processes of recovery, cleanup and quantification of DNA as they are applied to body fluid stains analyzed in the crime laboratory.

QUANTITY OF DNA IN HUMAN CELLS

Diploid human cells each contain about 6 pg of DNA, contained almost totally in the 46 chromosomes. About 1% of the total cellular DNA is located in the mitochondria. Haploid cells (spermatozoa and ova) contain one half the diploid quantity of DNA, or about 3 pg/cell. For human somatic cells, only the erythrocytes are devoid of chromosomal DNA.

Since crime scene evidentiary materials most often are shed body fluids, it is useful to estimate the quantities of DNA expected to be in common fluids. For example, the normal number of leukocytes in human peripheral blood generally ranges between 5 and 10 million cells per ml of blood. This equates to DNA concentrations that can vary between 30 μg/ml and 60 μg/ml of peripheral blood. The average number of spermatozoa per ml of semen is 150 x 10^6/ml (Mann and Lutwak-Mann 1981). Thus the DNA concentration in semen is considerably higher than in blood and averages about 450 μg/ml. Each ml of semen can contain about 5 x 10^6 leukocytes which contribute another 30 μg DNA per ml. Thus semen from an average male possesses approximately 480 μg DNA per ml.

Considering that a satisfactory autoradiogram image of a restriction fragment linked polymorphism (RFLP) pattern can be obtained from 20–50 ng of human genomic DNA, one can calculate that, with a 100% yield, the DNA would have to be recovered from 4,000–10,000 nucleated cells at a minimum. Since the yield of DNA from stain material routinely is less than 100%, more than the minimum number of nucleated cells is required to give a satisfactory RFLP pattern.

GOALS OF THE DNA RECOVERY PROCESS

There are three major goals that must be met during the DNA recovery processes if the RFLP typing analysis is to be successful. The recovered DNA must be of high molecular weight; the DNA must be free of substances that would compromise restriction endonuclease activity; and the quantity of DNA retrieved must be sufficient to meet the analytical requirements of the test procedure.

Factors Affecting the Size of Recovered DNA

There are several factors that can affect the molecular size of the DNA recovered from biological evidentiary specimens. One of the most common causes of DNA destruction is the action of the endo- and exonucleases that are ubiquitous in nature. While the analyst can do nothing to prevent the destructive action of nucleases prior to specimen collection, there are several steps that can be taken to abrogate their activity once the specimen environment can be controlled. Specimens should be maintained cool and dry prior to the commencement of DNA recovery procedures. The extraction solution should contain a metal chelator, such as ethylenediaminetetraacetic acid (EDTA), to prevent the activity of DNases. Moreover, pipette tips, glassware, and appropriate reagent solutions should be autoclaved before use to inactivate any nuclease activity that might be present.

A common abuse inflicted on DNA in the laboratory during recovery is excessive shear force, such as prolonged agitation on a vortex mixer. Such mishandling can result in DNA fragmentation to an extent that the molecular weights of the fragments are too low for the remaining steps in the RFLP procedure to be carried out.

The molecular weight of intact DNA in the cell nucleus is about 10^{11} daltons. Careful recovery of DNA that has not been severely affected by nuclease activity or shear forces should be of molecular weight about 10^7 daltons (Gross-Bellard *et al.* 1973).

Factors Affecting the Susceptibility of DNA to Restriction Digestion

The recovery process must not only avoid destruction of the DNA, but it must render the DNA suitable for digestion by restriction endonuclease (RE) activity. Two factors can reduce the ability of RE to fully digest DNA: the presence of histone proteins that remain attached to the DNA and pre-

vent access of RE to all possible restriction sites; and the presence of adventitious substances that are inhibitory to the catalytic activity of the RE.

Histones are removed from the DNA and denatured by the action of detergents. This is usually done at moderately elevated temperature (for example, 56° C). The denatured histones then are hydrolyzed to peptides and amino acids by the action of a proteolytic enzyme incorporated into the extraction solution.

Adventitious substances that are present in the DNA preparation can inhibit the catalytic activity of the RE. Such adventitious substances either were present on the surface upon which the body fluid was deposited (for example, dirt, salt, acids) or were introduced during the DNA recovery (for example, phenol, chloroform). Regardless of their source, contaminants can be removed by precipitation of the DNA from solution with ethyl alcohol, leaving the contaminants in solution; by dialysis; or by filtration.

Factors Affecting the Quantity of DNA Recovered

The quantity of DNA recoverable from forensic specimens is influenced by a number of factors. Obviously, the size of the stain or tissue sample affects the potential quantity of DNA that is available for recovery. In addition, the type of surface onto which the fluid has been deposited can dramatically affect the level of DNA recovery for body fluid stains.

GENERAL METHODS FOR RECOVERY AND EVALUATION OF DNA

The process of recovering DNA from forensic specimens can be broken into five steps: (1) Rehydration of the stain and solubilization of the stain components; (2) denaturation and hydrolysis of proteins; (3) removal of denatured proteins; (4) purification of the DNA; and (5) quantification of DNA and assessment of its quality for RFLP analysis. Each of these steps will be considered further.

Solubilization of Stain Components

Dried stains can lose considerable water in the course of drying. This water must be replaced and the stain components resolubilized for the recovery procedures to succeed. During this phase of the recovery process, the DNA must be protected from unnecessary degradation. Stains are cut and solubilization is accomplished by soaking the stain in buffer, usually at temperatures from 37° C to 56° C. A chelator of magnesium, such as EDTA, is included to prevent the action of the nucleases that

would destroy the DNA. Many procedures call for an overnight extraction period (Kanter *et al.* 1986; Gill *et al.* 1987).

Denaturation/Hydrolysis of Proteins

The presence of detergents in the stain extraction buffer is responsible for the lysis of cellular membranes and for the dissociation and denaturation of the histone proteins that are tightly attached to the DNA strands. Detergents destroy the secondary and tertiary structures of proteins which leads to their decreased solubility in aqueous solution and increased susceptibility to the hydrolytic activity of proteolytic enzymes. A commonly employed detergent is sodium dodecylsulfate (SDS).

Proteinase K (Ebeling *et al.* 1974) is widely used in procedures for the isolation of DNA as an effective tool for the hydrolysis of histone proteins. This enzyme is active across a wide range of pH, is active in the presence of SDS (in fact, its activity is enhanced), and it is unaffected by metal chelators such as EDTA.

Removal of Denaturation Products

Denatured proteins can be removed effectively from the extraction solution by treatment with phenol and chloroform. Phenol, and to some extent chloroform, is an effective protein denaturant. Moreover, the products of denaturation and proteolysis are soluble in phenol (Kirby 1957). Some recovery procedures call for the use of phenol and chloroform mixtures, while others use phenol first followed by one or more treatments with chloroform to ensure complete removal of phenol. Isoamyl alcohol is included in mixtures of phenol and chloroform to reduce the tendency of proteins to foam when they are denatured during shaking with the organic solvents (Marmur 1961).

Purification of DNA

Additional cleansing of the DNA is necessary before restriction digestion is attempted. The purpose of this step in the procedure is to remove small molecules that are potential inhibitors of RE catalytic activity. Three approaches to additional DNA cleanup that have been used are: (1) Precipitation of the DNA from solution by ethanol; (2) dialysis of the DNA solution against large volumes of buffer (Kanter *et al.* 1986); and (3) ultrafiltration through selectively permeable membranes (Marashi *et al.* 1985).

Assessment of Quantity and Quality of DNA

There are four approaches that can be used to determine the quantity or quality of DNA. While all four permit quantification of DNA, not all enable a comprehensive assessment of DNA quality.

Ultraviolet Absorption

The purine and pyrimidine bases in DNA absorb light in the ultraviolet (UV) region of the light spectrum. While each of the four bases that occur in DNA has its own specific absorption maximum, the composite maximum occurs at 260 nm. A DNA solution of concentration 50 μg/ml will yield an absorbency at 260 nm that is equal to 1.0. The lower limit of detection for DNA by UV is about 0.5 μg DNA/ml.

Ultraviolet measurements are useful also for determining if proteins and/or phenol remain with the DNA after isolation and purification. If proteins have been effectively removed from the DNA, the ratio of absorbencies determined at 260 nm and 280 nm will equal about 1.8. Likewise, when the ratio of absorbencies 260 nm/270 nm equals 1.2, the preparation can be considered to be free of phenol. If either protein or phenol remains in the preparation, the respective ratio of absorbencies will fall. However, UV determinations do not enable an estimate of the extent of DNA degradation.

Fluorescence Measurements

Without modification, DNA does not have the ability to fluoresce. Fluorescence measurements can be made only after a suitable dye has been allowed to interact with the DNA. Two dyes that have been used for this purpose are Hoechst 33258 (Brunk *et al.* 1979) and ethidium bromide. The excitation and emission maxima for both dyes shift when they are bound to DNA. Hoechst 33258 appears to bind preferentially to regions of DNA that are rich in A–T base pairs; whereas ethidium bromide intercalates between the stacked bases (Watson *et al.* 1987). Quantification of DNA by fluorescence is about 10 times more sensitive than by UV measurement. Fluorescence measurements only permit quantification of DNA, nothing can be learned about the molecular weight of DNA the level of contamination by protein or by phenol.

Yield Gel Measurements

The approximate concentration and molecular weight of DNA obtained from specimens can be readily determined by yield gel procedures. Yield gels are small agarose gels that contain ethidium bromide. A fraction of each DNA specimen along with DNA calibration standards are subjected to a brief electrophoretic separation at relatively high voltage. The standards cover the range from 15 ng DNA up to about 300 ng DNA, generally in doubling steps. After the electrophoretic run is complete, the gel is placed under UV light and a photograph is taken. For analyst safety, test gels should not be evaluated while the gel is irradiated with UV light. The intensity of fluorescence of the DNA test specimen(s) in the photograph is compared with the intensity of the standards and an estimate is made of the DNA concentration in the test specimen(s). This method of determining DNA concentration in test specimens is semi-quantitative and is the least accurate of the available quantitative methods.

Yield gel measurement is the most rapid method of assessing the molecular weight of the DNA in a test specimen. High molecular weight DNA remains as a compact band that does not migrate far from the origin during the brief electrophoretic period. In contrast, DNA that has been degraded will migrate more rapidly than high molecular weight DNA and if DNA has been degraded completely, the fluorescing band will migrate with or ahead of the bromophenol blue tracking dye that is added to each sample. Partially degraded DNA will be seen as a smear of fluorescing material that runs from the high to the low molecular weight regions of the gel. It is recommended that standards of RE digested viral DNA be included on each yield gel as comparative size standards. For example, RE digestion of lambda phage DNA by HindIII, yields six DNA fragments that range in size from 2027 base pairs (bp) up to 23130 bp. Yield gels do not enable estimates of protein or phenol contamination.

None of the quantitative/qualitative methods described to this point permit estimates of the human DNA that is present in a specimen. Since DNA other than human can potentially be present in a forensic specimen, these methods can result in an overestimate of the quantity of human DNA.

Slot/dot Blot and Human DNA Probe

This method, unlike those previously described, permits the specific estimation of the amount of human DNA in a specimen. To carry out the slot/dot blot test, DNA recovered from the specimen is denatured and bound to a nylon membrane. Generally, the membrane is held in an appa-

ratus that enables the application of multiple specimens on one membrane. The sample application area can have either the configuration of a dot or a slot, hence the term, slot blot or dot blot. The membrane-bound DNA specimens are then hybridized with a radioactively-labeled DNA probe that is complementary to a highly repetitive sequence found only in human DNA. An example of such a DNA probe is p17H8 (Waye *et al.* 1989) that is complementary to a highly repetitive sequence located in human chromosome 17. When this probe is used under low stringency conditions, it will hybridize with closely related sequences found in the DNA of other chromosomes. Because the concentration of the target sequences is high, many molecules of probe are bound and an intense autoradiographic signal can be obtained in just a few hours. Standards of DNA are placed onto the membrane along with the test specimens as quantitative references. This technique, although semi-quantitative, is extremely sensitive, enabling the detection of human DNA in the pg range. Dot/slot blot techniques do not inform the analyst about protein or phenol levels in the specimen, nor do they reveal anything about the intactness of the specimen DNA, unless the target DNA is so thoroughly degraded that the probe will not hybridize.

CONCLUSIONS

The isolation of DNA from forensic specimens is not a difficult procedure. As with any technique however, care must be exercised to insure that the limited quantities of DNA that are available in such specimens are not squandered by poor or sloppy analytical technique. It has been the experience of many individuals that improved yields of DNA are seen as analysts gain experience with, and confidence in, the procedures.

Newer methods for the recovery of DNA from forensic specimens may be on the horizon. Procedures have been published which show that DNA can be recovered effectively while avoiding the use of hazardous reagents such as phenol. These procedures utilize high NaCl concentrations (Miller *et al.* 1988), or chaotropic agents such as guanidinium hydrochloride or urea (Jeanpierre 1987; Lindblom and Holmlund 1988) to effect the removal of proteins from DNA.

REFERENCES

Brunk, C. F., Jones, K. C. and James, T. W. (1979). Assay for nanogram quantities of DNA in cellular homogenates, Anal. Biochem. 92:497–500.

Ebeling, W., Hennrich, N., Klockow, M., Metz, H., Orth, H. D. and Lang, H. (1974) Proteinase K from *Tritirachium album* limber, Eur. J. Biochem. 47:91–97.

Gill, P., Lygo, J. E., Fowler, S. J. and Werrett, D. J. (1987). An evaluation of DNA fingerprinting for forensic purposes, Electrophoresis 8:38–44.

Gross-Bellard, M. Oudet, P. and Chamron, P. (1973). Isolation of high-molecular-weight DNA from mammalian cells, Eur. J. Biochem. 36: 32–38.

Jeanpierre, M. (1987). A rapid method for the purification of DNA from blood, Nucleic Acids Res. 15:9611.

Kanter, E., Baird, M., Shaler, R. and Balazs, I. (1986). Analysis of restriction fragment length polymorphisms in deoxyribonucleic acid (DNA) recovered from dried bloodstains, J. Forensic Sci. 31:403–408.

Kirby, K. S. (1957). A new method for the isolation of deoxyribonucleic acids: Evidence on the nature of bonds between deoxyribonucleic acid and protein. Biochem. J. 66:495–504.

Lindblom, R. and Holmlund, G. (1988). Rapid DNA purification for restriction fragment length polymorphism analysis, Gene Anal. Techn. 5:97–101.

Mann, T. and Lutwak-Mann, C. (1981). Male Reproductive Function and Semen, Springer-Verlag, Berlin.

Marashi, F., Stein, G., Stein, J. and Schubert, C. (1985). Use of ultrafiltration microconcentrators in the concentration and desalting of DNA, Biotechniques May/June 238–240.

Marmur, J. (1961). A procedure for the isolation of deoxyribonucleic acid from microorganisms, J. Mol. Biol. 3:208–218.

Miller, S. A., Dykes, D. D. and Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells, Nucleic Acids Res. 16:1215.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. and Weiner, A. M. (1987). Molecular Biology of the Gene, 4th ed. The Benjamin/Cummings publishing company, Menlo Park, California.

Waye, J. S., Presley, L., Budowle, B., Shutler, G. G. and Fourney, R. M. (1989). A simple and sensitive method for quantifying human genomic DNA in forensic specimen extracts, Biotechniques 7:852–855.

RESTRICTION ENZYMES: BASIC PROPERTIES AND USE IN RFLP ANALYSIS

Bruce H. Howard

National Cancer Institute
Bethesda, Maryland

Forensic applications of restriction fragment length polymorphism (RFLP) analysis are increasing rapidly in both frequency and importance. Not surprisingly however, because RFLP results can have a decisive impact in certain types of cases, there is intense scrutiny concerning potential artifacts in this technology. To avoid such artifacts, it is essential that proper use of restriction endonucleases be followed and that appropriate controls are always included. The first goal of this chapter is to discuss fundamentals relating to the use of restriction enzymes. The second goal is to highlight potential pitfalls in the utilization of these enzymes that must be avoided if RFLP analysis is to become widely accepted.

DNA STRUCTURE AND RESTRICTION ENZYME RECOGNITION SEQUENCES

To aid in visualizing how restriction enzymes interact with DNA, a schematic drawing of the DNA double helix is presented in Figure 1. In this representation, the two phosphate sugar backbones are shown as ribbons wound around each other to form a twisted ladder. Between these two phosphate sugar backbones are the bases, which are paired by the rule that A matches with T and G matches with C. Restriction enzymes are thought to scan along the DNA helix searching for recognition sequences, that is, the sequences which they bind and cut. In this figure, brackets indicate the recognition sequence for EcoRI, one of the first restriction endonucleases discovered.

The characteristics of the prototypic EcoRI restriction endonuclease site are more clearly illustrated in Figure 2; here a further abstraction of DNA is depicted, with the helices untwisted and the phosphate linkages explicitly shown. It should be noted that the phosphate backbones are arranged in anti-parallel configuration. In the case of EcoRI, as is common for restriction enzymes, the recognition sequence is symmetrical, that is, the same sequence, GAATTC, is obtained if the DNA is read either on the upper strand from left to right or on the lower strand from right to left. Because the recognition site appears identical after it is subjected to a 180° rotation, it is said to possess dyad symmetry.

Once a restriction endonuclease binds to its recognition sequence, it cleaves the phosphate backbone. The arrows in Figure 2 show the points at which cleavage occurs for the enzyme EcoRI. Addition of water molecules to these points releases phosphate (PO_4) groups, which remain attached to the A residues, while hydroxyl (OH) groups are created on the G residues. Cleavage of the strands occurs sequentially, since time course experiments reveal initial nicking of one strand and then cutting of the second strand.

DNA METHYLATION

Restriction enzymes such as EcoRI, which have 6 base pair (bp) recognition sites, cleave DNA on average once every ≈ 3000 bp. Since these enzymes are active in the bacterial cells in which they are synthesized, and since the overall length of the bacterial chromosome is several million bp, the host chromosome would sustain extensive degradation if a protective mechanism did not exist. The protective mechanism that bacterial cells employ involves methylation. Methylases commonly occur pairwise with and have the same recognition sequences as the corresponding restriction endonucleases (such restriction enzyme-methylase doublets are termed cognate pairs). For example, the enzyme EcoRI methylase adds methyl groups on both strands to the second A residue in the sequence GAATTC, thereby rendering the recognition sequence resistant to cleavage by the cognate EcoRI restriction endonuclease (Figure 3). Newly synthesized DNA is hemi-methylated, that is, contains a methyl group on only one strand, but is comparatively resistant to cleavage and is rapidly methylated on the second strand.

Historically, the first evidence for paired restriction/modification systems derived from experiments in which researchers were attempting to transfer bacterial viruses (bacteriophage) grown in one bacterial strain to another strain. It was found that bacteriophage were often destroyed when introduced into the new strain. Further investigation revealed that in such instances bacteriophage DNA was degraded because it lacked methylation at appropriate sites to protect it from host nucleases.
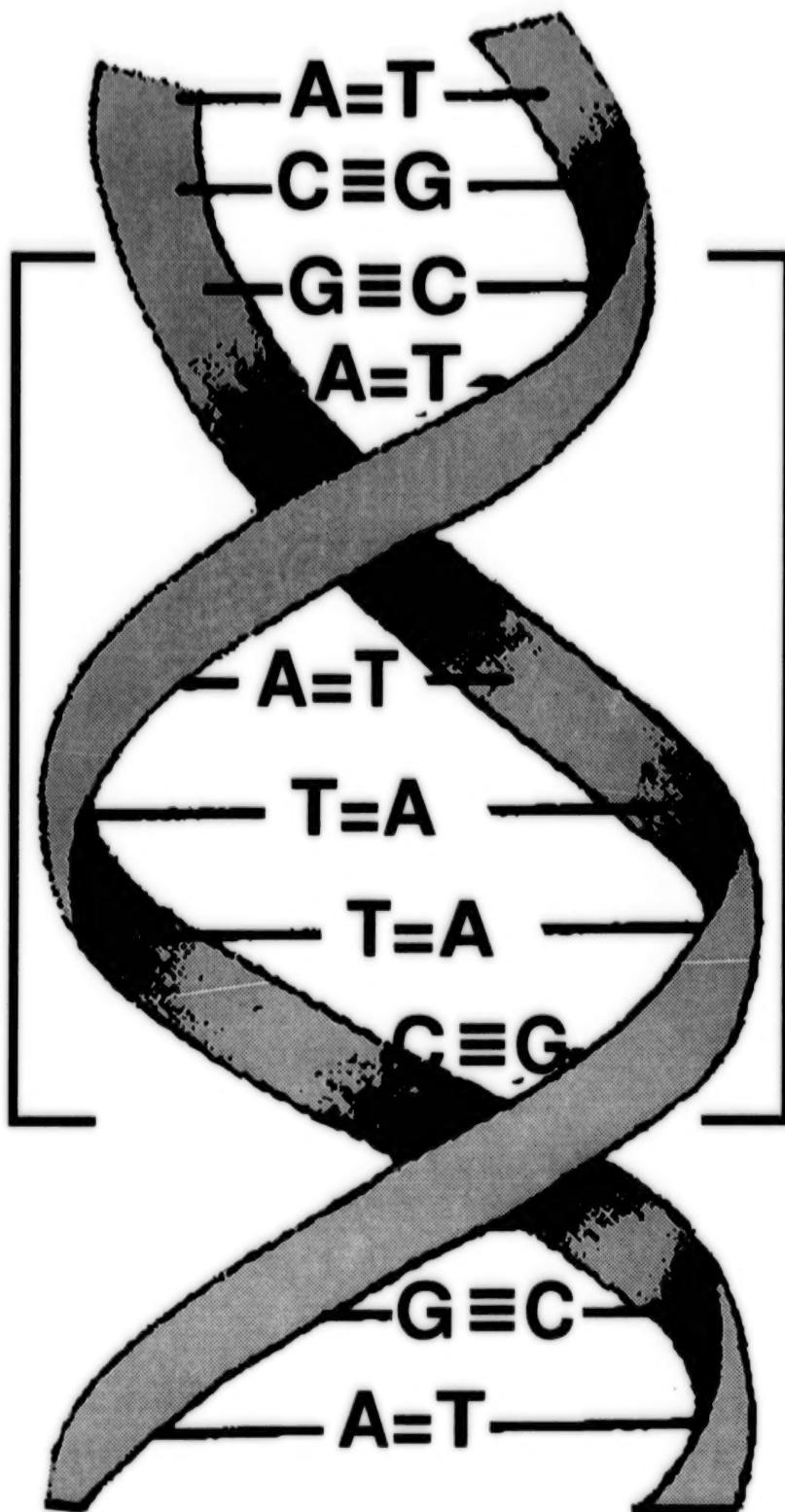
Figure 1. Schematic representation of DNA double helix. The deoxyribose-PO$_4$ backbones are represented as twisted ribbons. Base pairs are denoted by double or triple bars between bases A (adenine), C (cytosine), G (guanine), or T (thymine). Brackets indicate the extent of the recognition sequence (GAATTC) for the restriction endonuclease *Eco*RI.

These nucleases were said to restrict the transfer of bacteriophage between bacterial strains; hence the term restriction endonucleases.
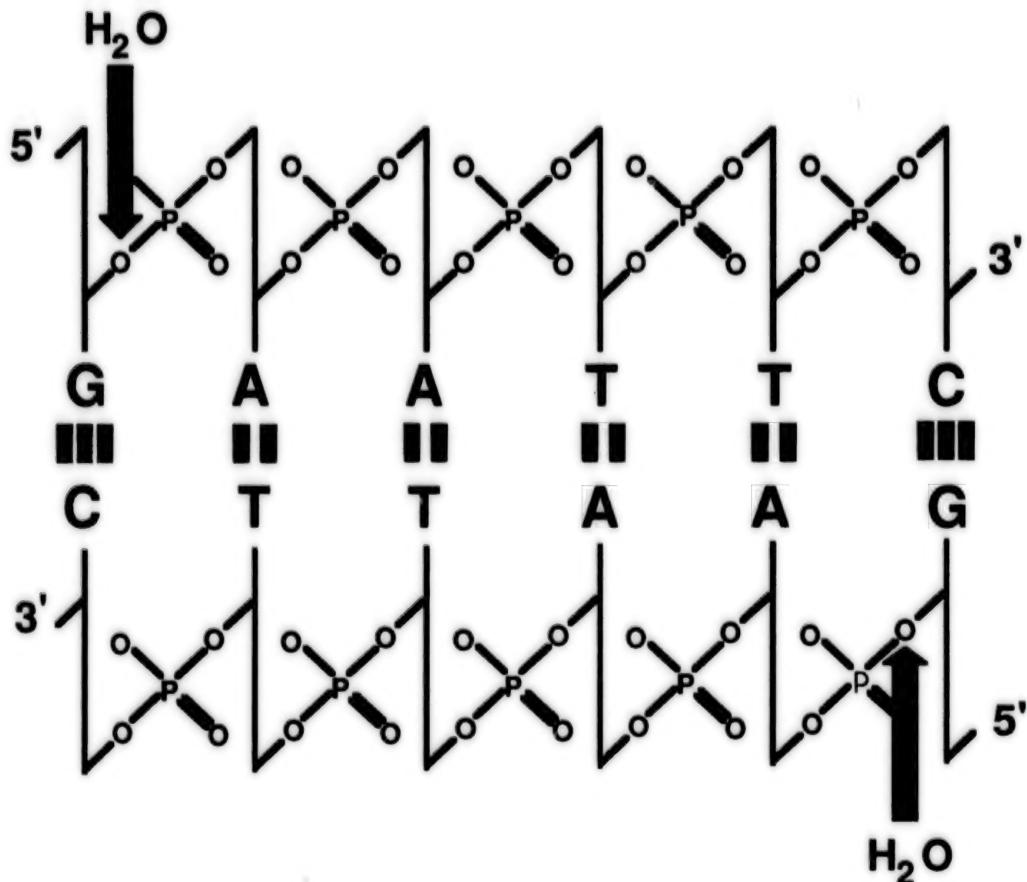
Figure 2. Cleavage at the *Eco*RI recognition sequence. In this abstract depiction of DNA, phosphate groups are explicitly shown, but deoxyribose residues are indicated only by vertical lines. Arrows indicate the positions at which addition of H_2O molecules results in cleavage of the phosphate backbones by *Eco*RI.

DIVERSITY OF RESTRICTION ENDONUCLEASES

The implication of the finding that restriction/modification systems could block bacteriophage transmission from one bacterial strain to another was that numerous different systems might exist. As it became apparent that restriction enzymes represented powerful tools for genetic engineering, a systematic search for new restriction enzymes ensued. This search revealed a remarkable diversity of restriction endonucleases (Roberts 1988). Sources include the Eubacteria and Archaebacteria as well as the fungus-like Streptomyces, commonly used for the production of antibiotics.

To date all restriction endonucleases discovered may be grouped into three types. Virtually all molecular cloning and forensic applications use type II enzymes; accordingly, examples and generalizations concerning restriction enzymes in this chapter refer exclusively to type II. Types I and III combine restriction and methylation activities within a single molecule and require adenosine tri-phosphate (ATP) for cleavage; type I enzymes cut randomly after interacting with their recognition sequence, whereas type III enzymes cut at a specific position (Maniatis *et al.* 1982).

Selected examples of restriction enzymes based on length of recognition sequence are listed in Table 1. The shortest sequence which is useful for RFLP analysis is four bp. This sequence will be recognized on the average of once every few hundred bp. Restriction enzymes with 4 bp recognition sequences are common, as are enzymes with 5 bp or 6 bp recognition sequences. There are also a few enzymes which cleave at longer recognition sites, for example, 7 bp or 8 bp. These latter enzymes are useful primarily for mapping extensive regions of the genomic DNA of mammalian cells.

Table 1 illustrates several other features of restriction recognition sites in addition to variations in length. First, restriction enzymes often tolerate ambiguity within their recognition sequences. The enzyme *Ppu*MI, for instance, allows occurrence at the central position of either A or T, and at the 5'

31
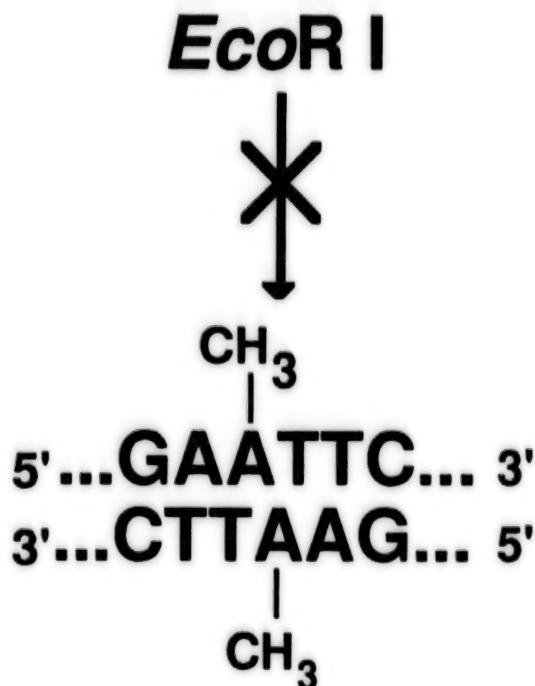
Figure 3. *Eco*RI methylase inhibition of *Eco*RI restriction endonuclease activity. *Eco*RI methylase adds methyl (CH₃) groups to A residues within the GAATTC recognition site. The modified site is completely resistant to cleavage by *Eco*RI restriction endonuclease.

and 3' edges of the recognition region a purine (A or G) and a pyrimidine (C or T), respectively. Obviously, although *Ppu*MI has a 7 bp recognition site, cleavage frequencies will reflect the sequence ambiguity that is allowed by this enzyme. A second feature of restriction endonuclease sites is that they may be hyphenated. At hyphenated sites the enzyme recognizes specific nucleotides on either side of a central group of unspecified nucleotides (usually indicated by the letter 'N'). In the example in Table 1, *Sfi*I cleaves at a hyphenated sequence consisting of two GGCC tetranucleotides flanking 5 unspecified nucleotides. Note that the enzyme *Hin*fI also allows an unspecified nucleotide within its cleavage sequence.

A final consideration with respect to diversity of restriction enzymes concerns position at which cleavage occurs within the recognition site. As shown by selected examples in Table 2, this position is variable. Enzymes such as *Eco*RI generate DNA fragments with short single strand ends that terminate with a 5' phosphate group; these are often referred to as 5' overhanging ends. In contrast, enzymes such as *Pst*I generate 3' overhanging ends, while other enzymes, for example, *Pvu*II, cleave within the center of the recognition site to create blunt ends. If an enzyme has a non-symmetrical recognition sequence, as in the case of *Fok*I, the cleavage site is commonly displaced relative to the recognition sequence.

Table 1. RESTRICTION ENDONUCLEASE RECOGNITION SEQUENCES

Length of recognition sequence	Example
4 bp *Rsa*I	5' . . . GTAC . . . 3'
5 bp *Hin*fI	5' . . . GANTC . . . 3'
6 bp *Hind*III	5' . . . AAGCTT . . . 3'
7 bp *Ppu*MI	5' . . . PuGGᴬ_TCCPy . . . 3'
8 bp *Sfi*I	5' . . . GGCCNNNNNGGCC . . . 3'

Examples are chosen to illustrate permitted ambiguities and variations in length of restriction enzyme recognition sequences. Pu, purine (A or G); Py, pyrimidine (C or T); N, any standard base (A, C, G, T).

Table 2. VARIATIONS IN CLEAVAGE SITES RELATIVE TO RECOGNITION SEQUENCES

*Eco*RI	5' . . . NNNG ↓ AATTCNNN . . . 3'3'	
	3 . . . NNNCTTAA ↑ GNNN . . . 5'	
*Pvu*II	5' . . . NNNCAG ↓ CTGNNN . . . 3'	
	3' . . . NNNGTC ↑ GACNNN . . . 5'	
*Pst*I	5' . . . NNNCTGCA ↓ GNNN . . . 3'	
	3' . . . NNNG ↑ ACGTCNNN . . . 5'	
*Fok*I	5' . . . NNNGGATGNNNNNNNNN	↓
	NNN . . . 3'	
	3' . . . NNNCCTACNNNNNNNNNNNNN	↑
	NNN . . . 5'	

Recognition sequences are indicated in bold letters with an arbitrary number of flanking unspecified bases (N). Positions of cleavage sites are designated by arrows.

32

RESTRICTION ENZYME DIGESTION CONDITIONS

Basic variables which affect the efficiency and specificity of restriction enzyme digestion (salt concentration, pH, temperature, etc.) can be critical to the outcome of an RFLP analysis. It is easy to neglect these seemingly trivial parameters; nevertheless, they are the most likely to cause aberrant cleavage patterns. For example, the salt (NaCl or KCl) concentration must be carefully checked, since the optimum concentration for digestion varies widely for different enzymes. Moreover, an inappropriate salt concentration can alter not only digestion rate but also cleavage specificity (see following discussion). Although less commonly a problem, incomplete digestion can also be caused by inappropriate pH (where pH is determined primarily by the Tris HCl buffer), or incorrect concentration of magnesium chloride. Most restriction endonucleases are incubated at 37°C during the DNA digestion; however, a number of enzymes originally derived from thermophilic organisms perform optimally at considerably higher temperatures (50–65°C) and a few enzymes give better results at lower temperatures (25–30°C).

Apart from basic variables in the restriction reaction, there are several variables that become important in specific situations, for example, with enzymes that are labile or for prolonged digestions. In such instances it is good practice to add freshly prepared reducing agent (for example, β-mercaptoethanol from a concentrated stock solution kept sealed at −20°C) and to include bovine serum albumin as a stabilizing agent.

CONTROLS FOR CORRECT RESTRICTION ENZYME DIGESTION

The second part of this chapter deals with the inclusion of controls to insure that restriction enzyme digestion has gone to completion and that no cleavage events have occurred at sites other than consensus recognition sequences. The standard method to confirm the accuracy of the digestion is to include in the analysis one or more reference DNA templates. By far the most reliable method is to perform an internal control reaction, that is, to mix the reference DNA template(s) in the same tube with the unknown DNA sample prior to initiation of the digestion. There are two categories of control DNA templates that may be used. Cloned DNAs with known patterns of restriction fragments represent the first category; such DNAs include bacterial plasmids, bacterio-

phage, or recombinant constructions containing the human DNA locus at which the RFLP of interest occurs. Human genomic DNA represents the second category: such DNA could be a standard reference DNA or a second unknown to which the first unknown is to be compared. Each of these categories of controls presents potential problems of which the investigator must be aware.

PARTIAL OR FAILED DIGESTION OF CLONED DNA DUE TO NON-COGNATE METHYLATION

Most *E. coli* strains commonly used for molecular cloning contain two methylases in addition to those that may be associated with cognate restriction endonucleases. These are the *dam* and *dcm* methylases, which modify GATC to G^{Me}ATC and CC^A_TGG to C^{Me}C^A_TGG, respectively (Hattman *et al.* 1978; Marinus and Morris 1973; May and Hattman 1975). Many restriction enzymes whose recognition sequences contain or overlap one of these methylation sequences will not cleave modified DNA (Nelson and McClelland 1987). For example, neither *Mbo*I (recognition sequence GATC) nor *Bcl*I (recognition sequence TGATCA) will cut *dam*-methylated DNA. A less obvious problem arises with the enzyme *Taq*I (recognition sequence TCGA), which fails to cleave where its site and a *dam*-methylation site overlap, that is, at TCG^{Me}ATC sequences. Such overlapping occurs on average 1 in every 16 *Taq*I sites. The failure of a restriction endonuclease to generate the number of fragments predicted by the primary DNA sequence of a cloned, control DNA can create confusion, leading to the erroneous conclusion that the same enzyme has also failed to completely cleave the human genomic DNA in an unknown sample.

PARTIAL DIGESTION OF MAMMALIAN DNA DUE TO 5′-METHYLCYTOSINE MODIFICATION

A more serious problem than prokaryotic methylation is the variable modification of human and most other higher eukaryotic DNAs at cytosine residues to form 5-methyldeoxycytosine (5mC). This modification occurs primarily at CG dinucleotides, where over 40% of cytosine residues may be in the 5mC form (Bird and Southern 1978; Woodcock *et al.* 1987). It also occurs at the dinucleotides CA, CC, and CT, albeit at a much lower frequency (<2%) (Woodcock *et al.* 1987). Formation of 5mL has the potential to confound RFLP analysis, because the level of 5mC at a particular

gene locus can fluctuate depending on the source of tissue. Thus, enzymes which are sensitive to this modification could generate dissimilar restriction patterns from two samples taken from different tissues of the same individual. Clearly this is unsatisfactory for forensic studies.

If it is necessary in a RFLP analysis to cleave at a site that is subject to 5mC modification, there are at least two ways in which this problem may be circumvented. In some cases, a restriction endonuclease is available that is not inhibited by the presence of 5mC within its recognition sequence. For example, *Msp*I will cleave $C^{M*}CGG$ sites, in contrast to its isoschizomer *Hpa*II (Ploeg and Flavell 1980; Nelson and McClelland 1987). When no modification-resistant isoschizomer exists, polymerase chain reaction (PCR) amplification can often be used to dilute out 5mC modification at the RFLP locus of interest prior to restriction enzyme digestion.

INCORRECT DIGESTION DUE TO STAR(*) ACTIVITY

Cleavage by a restriction endonuclease at a sequence similar but not identical to its defined recognition sequence is termed star activity. This type of abnormal activity usually reflects decreased specificity, that is, the enzyme allows greater ambiguity than normal in its recognition sequence, and most often occurs under inappropriate digestion conditions. Star activity-inducing conditions include lower than recommended ionic strength, high pH, presence of organic solvents such as glycerol or DMSO, or very high endonuclease concentrations (Chirikjian and George 1981; Gardner et al. 1982; Heininger *et al.* 1977; Hsu and Berg 1978; Malyguine *et al.* 1980; Nasri and Thomas 1986; Polisky *et al.* 1975; Tikchonenko *et al.* 1978; Woodbury *et al.* 1980).

There are several errors which may unmask star activity in a RFLP analysis. The first error is failure to adequately dilute a concentrated enzyme from its storage buffer; this will result in an excessive glycerol concentration in the restriction digestion reaction. The second error is addition of a massive excess of enzyme in an effort to avoid partial digestion. The third error is to perform the digestion at a lower salt concentration than recommended. In situations where two enzymes are used simultaneously but require different salt concentrations, the experimenter is frequently tempted to employ the lower salt concentration or an intermediate concentration; either of these choices may promote star activity. The correct procedure is to perform a first digestion with the low-salt enzyme, then add salt to raise the ionic strength and carry out digestion with the high-salt enzyme.

Obviously it is essential in forensic RFLP work to eliminate star activity, since the variable appearance of additional cleavage products invalidates results. Fortunately, the tendency to exhibit star activity varies widely for different restriction enzymes; thus by rigorously adhering to proper digestion conditions and choosing an enzyme that is not predisposed to star activity, it should be possible to avoid this pitfall.

VARIABLE DIGESTION RATES AT CONSENSUS RECOGNITION SITES

Restriction endonucleases are extremely valuable tools because, under defined digestion conditions, they cleave at virtually all consensus recognition sequences and cleave only at those sequences. However, the rates at which a given enzyme may cut different consensus sites commonly vary 10 to 15 fold (Brown and Smith 1977; Forsblum *et al.* 1976; Gingeras and Brooks 1983; Nath and Azzolina 1981; Thomas and Davis 1975). Moreover, in unusual cases >50 fold reductions in cleavage rate may occur (Schildkraut 1988). Although the mechanisms underlying variation in cleavage rate have not been fully explored, slow cutting at a particular site presumably is due most often to the presence of unfavorable nucleotide sequences immediately flanking the consensus recognition sequence.

In RFLP analysis the completeness of digestion is frequently determined by addition of control DNA that produces a known restriction pattern. This approach may be inadequate if the polymorphic locus to be tested contains a site which is cleaved more slowly than any of the sites in the control DNA. In such a case the control DNA may be digested to completion while the unknown polymorphic locus is only partially digested. Obviously, it is essential for the relative cleavage kinetics of both control and polymorphic sites to be well characterized.

A still more subtle RFLP analysis artifact that can be envisaged is the variable occurrence of a refractory site in an unknown sample as a result of cytosine methylation immediately adjacent to the consensus recognition sequence. The investigator may be alert to the possibility of incomplete digestion for enzymes which have CG dinucleotides in their recognition sequences, but not suspect a similar problem where a CG dinucleotide is adjacent to or overlaps a single nucleotide of the recognition sequences at a particular locus.

Finally, a secondary polymorphism, adjacent but unrelated to the primary polymorphism, could in principle generate an unanticipated resistance to digestion. As with the above mentioned hazards, it should be possible to eliminate this class of pitfalls by judicious choice of an enzyme that is relatively affected either by variations in sequences flanking its consensus recognition site or by flanking/overlapping 5mC residues.

REFERENCES

Bird, A. P. and Southern E. M. (1978). Use of restriction enzymes to study eukaryotic DNA methylation: I. The methylation pattern in ribosomal DNA from *Xenopus laevis*, J. Mol. Biol. 118:27–47.

Brown, N. L. and Smith, M. (1977). Cleavage specificity of the restriction endonuclease isolated from *Haemophilus gallinarum* (*Hga*I), Proc. Natl. Acad. Sci. U.S.A. 8:3213–3216.

Forsblom, S., Rigler, R., Ehrenberg, M. and Philipson, L. (1976). Kinetic studies on the cleavage of adenovirus DNA by restriction endonuclease *Eco*RI, Nucleic Acids Res. 3:3255–3269.

Gardner, R. C., Howarth, A. J., Messing, J. and Shepherd, R. J. (1982). Cloning and sequencing of restriction fragments generated by *Eco*RI*, DNA 1:109–115.

Gingeras, T. R. and Brooks, J. E. (1983). Cloned restriction/modification system from *Pseudomonas aeruginosa*, Proc. Natl. Acad. Sci. U.S.A. 80:402–406.

Hattman, S., Brooks, J. E. and Masurekar, M. (1978). Sequence specificity of the P1 modification methylase (M.Eco P1) and the DNA methylase (M.Eco *dam*) controlled by the *Escherichia coli dam* gene, J. Mol. Biol. 126:367–380.

Heininger, K., Hörz, W. and Zachau, H. G. (1977). Specificity of cleavage by restriction nuclease from *Bacillus subtilis*, Gene 1:291–303.

Hsu, M. and Berg, P. (1978) Altering the specificity of restriction endonuclease: effect of replacing Mg2+ with Mn2+, Biochemistry 17:131–138.

Malyguine, E., Vannier, P. and Yot, P. (1980). Alteration of the specificity of restriction endonucleases in the presence of organic solvents, Gene 8:163–177.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Marinus, M. G. and Morris, N. R. (1973). Isolation of deoxyribonucleic acid methylase mutants of *Escherichia coli* K-12, J. Bacteriol. 114:1143–1150.

May, M. S. and Hattman, S. (1975). Analysis of bacteriophage deoxyribonucleic acid sequences methylated by host- and R-factor-controlled enzymes, J. Bacteriol. 123:768–770.

Nasri, M. and Thomas D. (1986). Relaxation of recognition sequence of specific endonuclease *Hind*III, Nucleic Acids Res. 14:811–821.

Nelson, M. and McClelland, M. (1987). The effect of site-specific methylation on restriction-modification enzymes, Nucleic Acids Res. (Suppl.) 15:r219–r230.

van der Ploeg, L. H. and Flavell, R. A. (1980). DNA methylation in the human gamma delta beta-globin locus in erythroid and nonerythroid tissues, Cell 19:947–958.

Polisky, B., Greene, P., Garfin, D. E., McCarthy, B. J., Goodman, H. M. and Boyer, H. W. (1975). Specificity of substrate recognition by the *Eco*RI restriction endonuclease, Proc. Natl. Acad. Sci. U.S.A. 72:3310–3314.

Roberts, J. R. (1989). Restriction enzymes and their isoschizomers, Nucleic Acids Res. (Suppl.) 17:347–387.

Schildkraut, I. (1988). Site preferences of restriction endonucleases. In: New England Biolabs Catalog, Pg. 133.

Thomas, M. and Davis, R. W. (1975). Studies on the cleavage of bacteriophage lambda DNA with *Eco*RI Restriction endonuclease, J. Mol. Biol. 91:315–328.

Tikchonenko, T. I., Karamov, E. V., Zavizion, B. A. and Naroditsky, B. S. (1978). *Eco*RI activity: enzyme modification or activation of accompanying endonuclease? Gene 4:195–212.

Woodbury, C. P. Jr., Hagenbüchle, O. and von Hippel, P. H. (1980). DNA site recognition and reduced specificity of the *Eco*RI endonuclease, J. Biol. Chem. 255:11534–11548.

Woodcock, D. M., Crowther, P. J. and Diver, W. P. (1987). The majority of methylated deoxycytidines in human DNA are not in the CpG dinucleotide, Biochem. Biophys. Res. Commun. 145:888–894.

BLANK PAGE

GEL ELECTROPHORESIS OF DNA

Lawrence I. Grossman

Wayne State University School of Medicine
Detroit, Michigan

The movement of DNA through a solid support in an electrical field, gel electrophoresis, has been an integral tool of the recombinant DNA decade. The major original use of analytical size determination has been augmented more recently by preparative separations of the components of complex mixtures and conformational studies on closed circular DNA utilizing one- and two-dimensional gels. Also, more specialized studies have been made possible by the great sensitivity of gels to small changes in conformation and mass; the studies discussed below deal with bending in linear DNA, and with band shift experiments, which allow the identification of specific DNA binding proteins in cellular extracts by assaying the mobility change of a protein-DNA complex compared to free DNA. Overall, the major advantages of gel electrophoresis are its enormous resolving power, the simple and relatively inexpensive apparatus utilized, and its easy combination with nucleic acid hybridization (blot transfer) methods, which assay for identity or near identity of DNA sequence. These factors have allowed electrophoresis to virtually replace ultracentrifuge sedimentation velocity, the main mass transport method in previous use.

In favorable circumstances DNA electrophoresis is capable of resolving DNA fragments differing by a single nucleotide, amounting to as little as 0.1% in molecular weight, or by a single superhelical turn in molecules of identical molecular weight. Nevertheless, its applications and its potential still appear to be increasing. One of the fastest growing directions is with respect to methods for periodically altering the direction of the field during electrophoresis to take advantage of the fact that reorientation of DNA in an electrical field is a stronger function of DNA size than is uniform movement. These pulsed field technologies have allowed the resolution and size determination of DNA molecules between 1 and 10 million bases, an improvement of one to two orders of magnitude over conventional separations in uniform fields. Each advance in separation technology, furthermore, can be coupled with blot transfer hybridization.

In this brief review I will discuss the basic principles and applications of DNA electrophoresis, and then highlight several enhancements that provide particular amplification of its power.

BASIC PRINCIPLES

Migration of DNA in Gels

DNA is a polyelectrolyte with a full negative charge at neutral Ph carried by the phosphates that form part of its backbone. As such, it will migrate toward the cathode in an applied electric field. The supporting gel commonly employed is unnecessary for electrophoretic movement, which in liquid is termed free electrophoresis. However, the gel serves at least two important functions. One is to stabilize migration against thermal and mechanical convection, and the other is to introduce a component of gel filtration chromatography. Since the charge density of DNA is independent of size, free electrophoretic mobility should also be independent of size (and shape), and this has been found (Olivera *et al.* 1964; Serwer and Allen 1984). However, virtually all applications of DNA electrophoresis depend on being able to separate molecules that differ in size and/or shape. The size- and shape-dependent mobility commonly observed is caused by gel sieving, in which the electric field functions as the driving force in a chromatographic type of separation.

Gels are almost always cast of agarose, a polysaccharide subfraction of agar, an algal extract, or of polyacrylamide. The choice generally depends on the size of DNA to be examined. Acrylamide gels are useful in a size range from several nucleotides up to several thousand, whereas DNA in agarose gels is well-behaved in a size range between 100 base pairs (bp) and 20–30 kilobase pairs (Kb) under conditions of a continuously applied field, and up to a substantially larger size, as will be discussed below, when the field is manipulated in various ways.

Most often, DNA is detected by staining with ethidium bromide and detecting the enhancement of its fluorescence when bound to DNA, or by autoradiography. As little as perhaps 10 ng can be detected by the former and at least an order of magnitude less by the latter.

Resolving Power

Under appropriate conditions DNA molecules in the range of less than 1 Kb that differ by a single nucleotide can be resolved from each other. Resolution in this range is routinely applied to DNA sequence analysis. However, substantially larger DNA molecules can be separated from each other on the basis of modest size or conformational differences. The actual size difference that can be resolved depends on the size range under consideration and the electrophoresis conditions. Gel concentration, and to a lesser extent field strength, are of particular importance in setting the upper limit for useful separation. For instance, the limit of well-behaved migration in 1.5% agarose, about 6 Kb, can be extended to at least 26 Kb at 0.3% agarose (Helling *et al.* 1974). This will be discussed further in the next section.

Comparison with the previous transport method of choice—sedimentation velocity—is particularly revealing. In this example (Johnson *et al.* 1979) replicative forms (RF) of bacteriophage φX174 DNAs are separated by sedimentation velocity (Figure 1B) into the traditional conformational forms, supercoiled closed circles (form I) and nicked circles (form II). Both form approximately Gaussian bands, although some overlap is clearly visible. Electrophoretic analysis of selected fractions, by contrast, shows a larger array of species, with no visible overlap. Particularly interesting is fraction 20, a local minimum in the sedimentation that does not return to baseline but appears composed of overlapping Gaussians. By gel analysis, in fact, this fraction contains significant amounts of both RF species. In addition to the major species, complex forms [circular dimers that are closed (DiI) and nicked (DiII)] accompany DNA I and linear DNA (form III) is present in the DNA II band.

Size Dependence

The most common use of DNA electrophoresis is determination of DNA size. As noted, the size-dependent mobility commonly observed is caused by electric field-driven gel sieving. Thus, it is not surprising that, for DNA that is not too large, DNA mobility is proportional to size. The nature of the proportionality has been examined empirically and under many conditions it is found that the log of size is related approximately linearly to mobility in gels of agarose or polyacrylamide (for example, Helling *et al.* 1974; Maniatis *et al.* 1975). Other relations have also been introduced
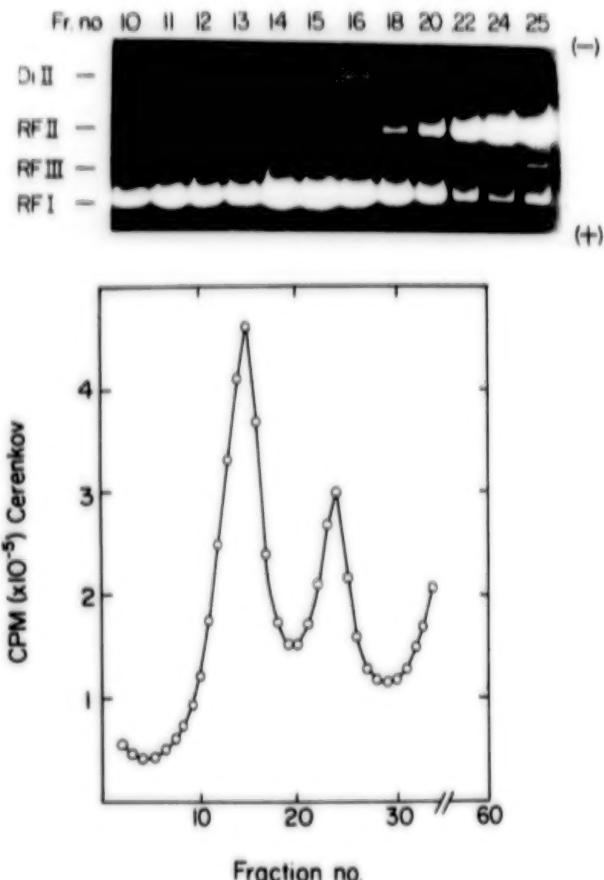
Figure 1. Comparison of sedimentation velocity centrifugation and agarose gel electrophoresis. (B) Gradient sedimentation. 32P-labeled DX174 RF DNA was sedimented through a 5-20% sucrose gradient in 50 mM Tris-HCl (pH 8.0), 5 mM EDTA, 50 mM NaCl. Centrifugation was performed in a Spinco SW27 rotor at 10′, 23,000 rpm, 20 hr. Fractions (0.5 ml) were pumped through a capillary inserted through the gradient and Cerenkov radiation determined. The field is directed toward the left. (A) Gel electrophoresis. Aliquots of the indicated gradient fractions containing approximately equal radioactivity were separated by electrophoresis on a 1% agarose slab gel at 3.3 V/cm, 5 hr. Electrophoresis took place in 40 mM Tris base, 5 mM sodium acetate, 1 mM EDTA (pH 7.8).

(Southern 1979). These relations allow DNA size to be evaluated relative to standards either graphically or by calculation, including by the use of a cubic spline function (Gariepy *et al.* 1986) to constrain the fitted curve to pass through the experimental points.

Deviations from linearity on a log (size) versus mobility graph as DNA size is increased occur in practice. At a constant field strength, the size at which linearity is lost depends on gel concentration. The illustration for agarose gels (Figure 2) (Helling *et al.* 1974) shows that an approximately tenfold extension of the linear range can be achieved by reducing the gel concentration from 1.5% to 0.3%. The fragility of gels at the lower

concentration precludes the use of lower concentrations in most instances.
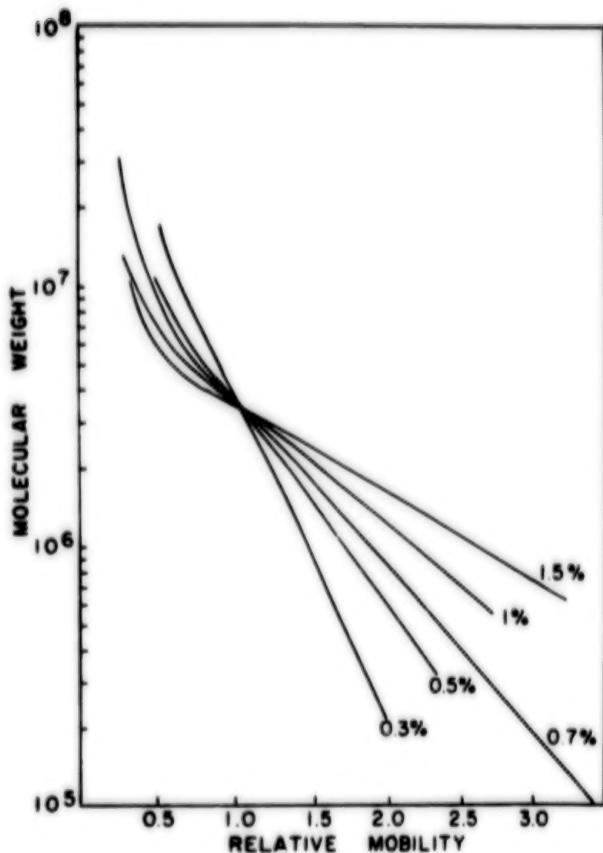
Figure 2. Relative electrophoretic mobilities of restriction endonuclease digests as a function of agarose concentration. From Helling *et al.* (1974).

The field strength also affects the effective size range for duplex DNA. This arises because DNA mobility, unlike that of proteins, is affected by field strength (Fisher and Dingman 1971; Flint and Harrington 1972; Johnson and Grossman 1977) (Table 1) and the magnitude of the effect is a nonlinear function of DNA size (Maniatis *et al.* 1975; Fisher and Dingman 1971; Table 1). The mechanism that has been proposed is the voltage gradient-induced deformation of random coils (reviewed in Serwer 1983; see also Deutsch 1988). The deviations from linearity become so marked that DNAs greater than about 20 Kb have a mobility at high fields that is independent of size. A computer simulation of a model (Deutsch and Madden 1989) that considers DNA to be a chain of charged beads negotiating a rectangular array of obstructions, the gel matrix, has received support from experiments in which DNA migration is directly visualized by attaching fluorescent groups to DNA that is migrating through a gel on a microscope slide (Smith *et al.* 1989; Schwartz and Koval 1989).

Table 1. EFFECT OF FIELD STRENGTH ON DNA MOBILITY

DNA	Mobility increase		
	0.6% T	1.0% T	1.4% T
DX174			
I	1.5	1.5	1.5
II	1.7	1.9	2.0
III	1.8	2.2	2.3
PM2			
I	1.6	1.8	1.8
II	2.3	3.7	3.7
III	2.1	3.1	3.5
MtDNA			
I	1.9	2.6	3.3
II	3.3	6.8	8.1
III	2.3	3.6	5.3

Although DNA deformation normally limits the range of effective size-dependent migration to about 15–25 Kb, some flexibility is possible by reducing the gel concentration and/or the field strength (Helling *et al.* 1974; Serwer 1980) to gain as much as an order of magnitude. Additional gains to the range of fungal chromosome-sized molecules and beyond have been accomplished by taking advantage of the size-dependent reorientation of DNA (Schwartz and Cantor 1984; Carle *et al.* 1986; Vollrath and Davis 1987; Southern *et al.* 1987; Serwer 1987), which is treated in the following discussion.

Effect of Loading Mass

A typical DNA blot transfer (Southern) hybridization (see below) contains lanes with genomic DNA digested with one or more restriction endonucleases, typically 5 to 10 μg, and lanes of size standards, typically 0.05 μg or less. Sometimes ignored is the effect of initial loading mass on migration, which has been demonstrated both for RNA (Richards and Lecanidou 1971) and DNA (Myers *et al.* 1976; Johnson *et al.* 1980). This effect can be substantial, amounting in the example shown (Figure 3) to a 55% increase in the center of mass of labelled DNA II when the amount of unlabelled DNA is increased from 0 to 5 μg. The major effect will be on larger molecules.
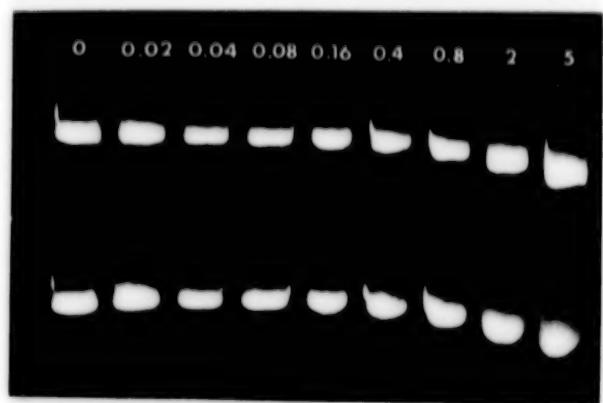
Figure 3. Effects of added DNA on conformational forms of bacteriophage PM2 DNA. Top: Autoradiogram of an 0.7% agarose gel run at 1.9 V/cm, 15 hr. Each lane contains 32P-labelled PM2 DNA (<0.05 Mg) containing equal amounts of form I (leading band) and form II DNAs and a minor amount of form III. The indicated amount (Mg) of unlabelled PM2 DNA (90% form I) is also present. Bottom: Densitometer tracing of lanes 0 and 5. The distance between forms I and II in lane 0 is 2.5 cm.

EFFECTS OF DNA CONFORMATION

Curved Linear DNA

Some DNA contains natural bends. Such sequence-directed curvature of DNA, which is now well-established, impinges on electrophoresis in two ways. First, such fragments migrate more slowly than their size would predict. Indeed, bent DNA was first identified in the kinetoplast DNA of a trypanosome by the aberrant electrophoretic

behavior of a fragment containing a bend (Marini et al. 1982). Although a special case, bent DNA provides another example of the failure of linearity between log (size) and mobility: a fragment containing a bend decreases in mobility compared to a linear fragment of the same size.

The other side of the coin is that the failure of the size-mobility relation can be utilized to map the position of a known or suspected bend (Wu and Crothers 1984). This is done by comparing the gel mobility of a set of fragments that are circularly permuted; that is, the suspected bend is put into a circular structure, restriction enzymes are used to open the circle in different places, and the mobility of the set of identically sized fragments is compared. This approach takes advantage of the fact that a bend near the end of a linear molecule (pole with hook on end) has less affect than a bend near the middle (boomerang). The magnitude of the effect was recently illustrated with SV40 DNA (Hsieh and Griffith 1988) by observing the migration of three fragments containing a curvature of about 200° near their centers.

The discovery of natural bending by electrophoretic analysis raises the question of whether it has a biological function. Although a definitive answer is not available, it is now known that sites of natural bending interact with specific proteins in several systems investigated. Furthermore, a nicking enzyme purified from Crithidia fasciculata reversibly decatenates kinetoplast DNA networks and specifically binds at the bend in this DNA (Linial and Shlomai 1987); the binding was shown to depend on bending rather than sequence. In some cases, binding enhances bending, such as at a plasmid (Koepsel and Khan 1986) or bacteriophage (Zahn and Blattner 1985) replication origin.

Dependence on Conformation

Under nondenaturing conditions complementary single strands sometimes can be separated electrophoretically from each other (and from the duplex from which they arose) (for example, Johnson and Grossman 1977). The separation arises because the sequence difference of the complementary strands allows different intrastrand pairings and thus conformational differences. At the other extreme, gels that are denaturing due to the presence of agents like urea, methylmercury (Bailey and Davidson 1976) or alkali (McDonnell et al. 1977) allow separation on the basis only of size. An interesting intermediate application is the use of a gel containing a gradient of denaturing power to detect subtle differences in the earliest melting

region between otherwise identical molecules (Lerman *et al.* 1984). This approach has been utilized to detect single base pair differences (Myers *et al.* 1985; Noll and Collins 1987).

Perhaps the major interest in conformational effects in electrophoresis, however, has been in the behavior of conformational isomers and in superhelical variants. The migration of DNA molecules depends upon both their size and shape. The effect of shape can be seen most clearly by holding size constant and considering the conformational interconversions of closed circular (supercoiled) DNA (form I), which can be nicked in one strand to generate a nicked circle (form II) or in both strands to generate a linear species (form III).

For DNA of moderate size analyzed under common electrophoretic conditions, form I will migrate most rapidly, followed by forms III and II. Figure 4 summarizes in diagrammatic form results for DNAs ranging from 5 to 16 Kb, agarose gels from 0.6 to 1.4%, and field strengths from 1.6 to 6.6 V/cm (Johnson and Grossman 1977). These DNA sizes are typical of plasmids, many viruses, recombinant molecules and mammalian mitochondrial DNA. The effect of gel concentration is seen to be largest on form III; for instance, for PM2 DNA at all field strengths above 1.64 V/cm, the order of forms I and III reverses as the gel concentration is raised. The effect of field strength at constant gel concentration is more complex since it reflects not only the different effective radii in the absence of a field but also the different susceptibility of each form to voltage-induced deformation. In any case, field strength effects are lowest for form I over the entire size range examined, and consequently cause reversals of migration order of some species. Thus, assignment of conformational forms based on migration order alone is premature.

Superhelical DNA

The migration of form I DNA is itself complicated by the fact that mobility is affected by the number of superhelical turns. DNA containing zero turns under electrophoresis conditions (relaxed DNA) will co-migrate with nicked circular DNA whereas supercoiled DNA as found in nature migrates more rapidly than nicked or relaxed circles. This wide variation in the electrophoretic behavior of closed circular DNA requires a brief background discussion.

In double-stranded DNA in solution, one strand winds around the other with a helical repeat of 10.4 bp (Wang 1979). There are numerous occasions when one strand winds or unwinds with re-

spect to the other, such as in denaturation, replication, or in the binding of some proteins or other ligands. In a closed circle, where there is no swivel point and therefore a topological constraint is present, the DNA must respond to the strain of change in duplex winding by rotating on itself, producing supercoils or twisted circles. All circular DNA found in nature is supercoiled to varying degrees; since a supercoil is a more compact structure than a circular ring, and, over a moderate range, the compactness increases with the number of supercoils, it is not surprising that there is an effect of DNA supercoiling on electrophoretic mobility.

Three parameters are needed to describe the topology of circular DNA molecules. The first is an invariant integer that counts the number of rotations of one strand about the other, now called the linking number (α or Lk). As originally formulated (Vinograd *et al.* 1968) (with the nomenclature updated—reviewed in Vosberg 1985), in a molecule containing no superturns, $\alpha = \alpha_0$, where α_0 is the number of duplex turns, calculated as N, the total bp, divided by 10.4, the average number of bp per duplex turn (Wang 1979). DNAs found in nature are always underwound, so that $\alpha < \alpha_0$, and $\alpha - \alpha_0$ is defined as $\Delta\alpha$ (originally τ), the number of superhelical turns that were topologically necessary to compensate the underwinding; for SV40 DNA, for instance, it is about 5% the number of duplex turns. The length-independent expression for the number of superturns, the superhelix density, is $(\alpha - \alpha_0 T40)\alpha_0$. DNAs found in nature vary about threefold in superhelix density. Additional insights have led to the modification of several concepts in the initial descriptions of supercoiling. The number of superhelical turns, although capable of experimental measurement, was realized to be conceptually deficient, since it applies to conformational rather than topological properties. Thus, the linking difference, $\Delta\alpha/\alpha_0$, is used to compare size-independent topological properties of DNA. The number of superhelical turns is thus a titrable subset (for example, with ethidium bromide) of the linking difference (Bauer 1978; Crick 1976; Crick *et al.* 1979).

A second modification has been the discovery that the value of A for a DNA sample is ordinarily not a single integer but a distribution centered about one or several dominant species. In fact, the distribution has been shown (Pulleyblank *et al.* 1975) to be a Boltzmann distribution of topological isomers. Since the free energy of supercoiling for introducing a small number of turns (1 to 3) at 37° into a DNA of about 5 Kb is 0.1 to 1 kcal/mole—

on the order of the energy of thermal motion at that temperature (0.6 kcal/mole)—it is not unexpected, in retrospect, that appropriately treated DNA (that is, with a topoisomerase) will come to thermal equilibrium as a distribution (Pulleyblank *et al*. 1975; Depew and Wang 1975). Indeed, a single band from a distribution excised from a gel and treated with a topoisomerase will regenerate the entire distribution (Pulleyblank *et al*. 1975).

Gel electrophoresis has been the major tool for deducing the above properties of supercoiled DNA, and of measuring the number of superhelical turns. Gels were used for identification of topoisomerase intermediates (Keller and Wendel 1974; Keller 1975a) and for counting the total number of titrable superhelical turns (Keller 1975b; Shure and Vinograd 1975). However, in gels such as that represented in Figure 4 supercoiled DNA migrates as a single band. The reason is that DNAs found in nature have too many superturns for a difference of one or several to alter the conformation enough to affect mobility. When most of these supercoils are removed, so that a DNA might contain a set of topological isomers with an average of about five, then each change of $VA = 1$ produces a conformational change that is visible as a separate electrophoretic band. Reversible removal of superturns is accomplished by titrating with planar polycyclic hydrocarbons such as ethidium bromide (EB), which intercalates between bases and thereby un-

winds the helix. Each EB bound unwinds the helix 26°.

Although one-dimensional gels do not ordinarily separate DNAs containing both positively and negatively supercoiled DNA of equal numbers of titrable superhelical turns, two-dimensional gels can do so (Lee *et al*. 1981). In such gels, the first dimension spreads out DNA normally so that supercoiled DNA (of both signs) migrates faster than relaxed or nicked circular species. In the second dimension at right angles, a low concentration of ethidium bromide or chloroquine (Shure *et al*. 1977) shifts the distribution of DNA so that originally negatively supercoiled DNA partially unwinds and thus migrates relatively slowly and DNA near relaxation or containing positive supercoils gains positive supercoils and migrates relatively more rapidly, thus leading to the typical arch pattern (Lee *et al*. 1981). Chloroquine binds less tightly to DNA than ethidium bromide, and thus its concentration is not as critical.

EXTENSIONS OF ELECTROPHORESIS

Blot Transfers

The power of electrophoresis to separate DNA fragments, particularly those generated by site-specific (restriction) endonucleases, can be enhanced significantly by coupling separation to the detection of specific fragments. In this extension,
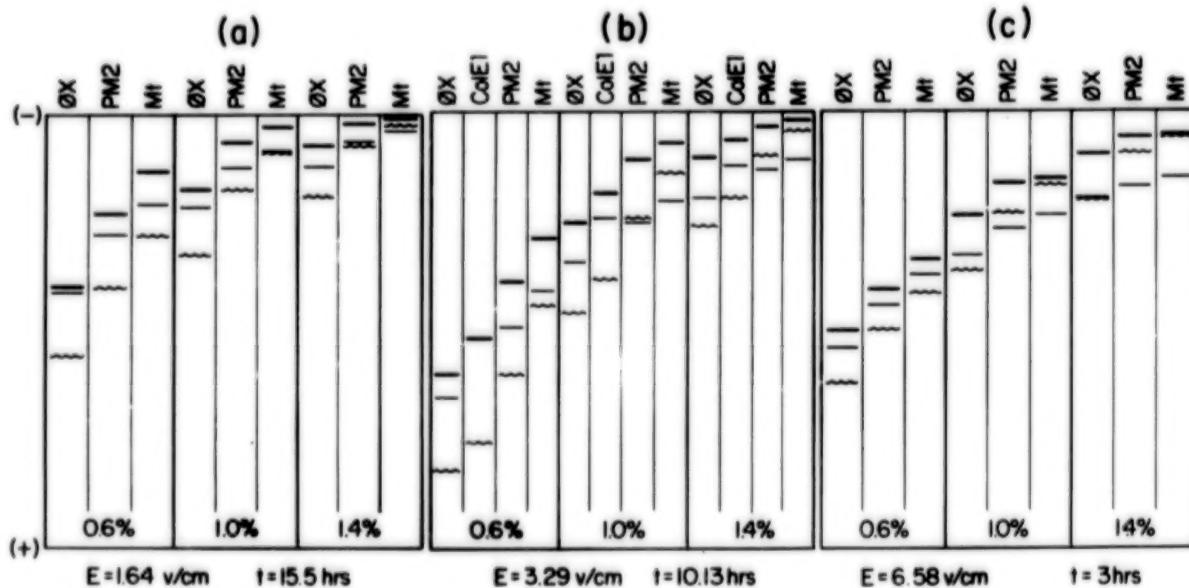
Figure 4. Diagrams of electrophoresis of DNA conformational isomers under various experimental conditions. DNA sizes (in Kb) are: DX174, 5.4; colE1, 6.4; PM2, 10.1; mouse mtDNA, 16.3. Agarose gels at the gel concentrations and voltage gradients shown were run in 40 mM Tris base, 5 mM sodium acetate, 1 mM EDTA (pH 8.2). The diagrams represent 15-cm long gels. Symbols: form I ($\sim$), form II (ZZZ), form III (———).

the DNA fragments in a gel are transferred after electrophoresis to a membrane (Southern transfer), and a particular fragment of interest can be detected by hybridization with a probe. An illustration (Figure 5) shows the assignment of coding sequences for the small ribosomal RNA to a particular restriction fragment of mitochondrial DNA from the water mold *Achlya ambisexualis.*

In the original method (Southern 1975, 1979), DNA was denatured in the gel and transferred by capillary flow of buffer to a nitrocellulose membrane, which retains the original pattern. A number of technical variations have since been introduced at almost all stages. For instance, the membrane may be made of nylon or polyester-reinforced nitrocellulose, both sturdier than the original; transfer may be carried out electrophoretically or by vacuum; and detection of a hybridized probe by its radioactive signal is no longer the only option. The most important variations, however, have been in extending the range of materials examined. RNA (Northern blot) and protein (Western) may be similarly separated by gel electrophoresis for blot transfer detection of specific bands. For RNA this detection is by hybridization, as for DNA, whereas for protein it is by reaction with a labelled antibody.

Gel Retardation

The mobility of a DNA fragment is reduced when protein is bound. This observation has been utilized to assay the specific binding of proteins (Fried and Crothers 1981), even in crude extracts (Strauss and Varshavsky 1984). In practice, 150–200 bp fragments of labelled DNA capable of putative specific binding is mixed with protein and an excess of non-specific DNA and subjected to electrophoresis at low ionic strength. In the case of DNA fragments containing a bend caused or enhanced by protein binding, gel retardation may depend both on the increased mass of the complex and on the increased bending of the DNA. A review of the technique and a discussion of applications and novel approaches has recently appeared (Revzin 1989).

PULSED FIELD ELECTROPHORESIS

The size limit discussed above for separating linear DNAs as a function of size pertains to DNA moving in a uniform field. Workers seeking a new way to separate DNA molecules on the basis of size in an electric field built upon the work of Zimm and his collaborators on the viscoelastic relaxation time of DNA. When DNA is distorted by

Figure 5. Blot transfer hybridization of Achlya mitochondrial DNA digested with BglII and probed with yeast mitochondrial 15S rDNA. The gel was stained with ethidium bromide (S), transferred to nitrocellulose, and hybridized at 50' with 32P-labelled 15S DNA, and autoradiographed (A). From Hudspeth *et al.* (1983).

a shear gradient, the time for return to the relaxed state depends upon the 5/3 power of molecular weight (Klotz and Zimm 1972). To harness this property for electrophoretic separations, Schwartz and Cantor (1984) utilized two alternating electric fields. Since each is on for only part of the run, this technology is called pulsed field gel electrophoresis. When the first field is applied, DNA elongates in the direction of the field. When this field is removed and another activated at some angle to the first, the DNA must first reorient before it can change direction. It is this reorientation time that is sensitive to size.

The original implementation of pulsed field electrophoresis, although it was groundbreaking and immediately useful, has stimulated many variations to eliminate some of the shortcomings. Nearly a dozen variations have since been introduced, including field inversion (FIGE; Carle *et al.* 1986), in which a uniform electric field is periodically inverted; CHEF (Chu *et al.* 1986; Vollrath and Davis 1987), in which homogeneous electric fields are produced with multiple electrodes used around a closed contour; and rotating gel electrophoresis (Serwer 1987; Southern *et al.* 1987), in which a uniform field is applied to a gel that is periodically and discontinuously rotated. Most recently, PACE (Lai *et al.* 1989) allows independent regulation of the voltages on 24 electrodes arranged in a closed contour.

The use of pulsed field techniques allows separation of DNA molecules up to about 6,000 Kb (6 mb). This range includes the sizes of yeast (*Saccharomyces cerevisiae*) chromosomes (0.2–3 mb), whose separation is illustrated (Figure 6). One application of such separations, in conjunction with Southern transfers, is the direct chromosome location of any gene for which a probe is available. Although current limits are still several orders too low to fractionate human chromosomes, improvements in gel electrophoresis, as well as other developments that are beyond the scope of this article, such as the use of rare cutting restriction enzymes and the construction of artificial yeast chromosomes with human or other DNAs, are helping to close this gap.
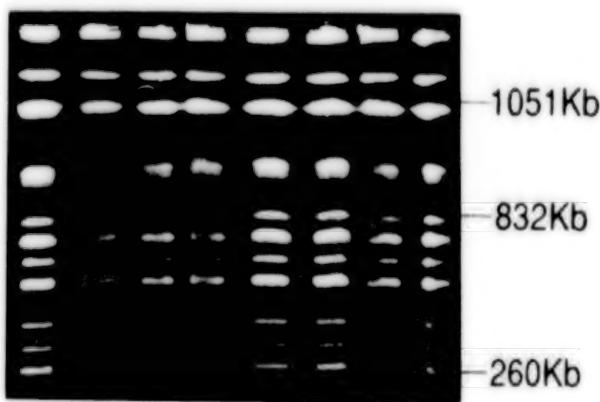
Figure 6. Pulsed field gel separation of DNA from chromosomes of Saccharomyces cerevisiae.

The effect of field strength on electrophoretic mobility (M) of DNA is illustrated with data from Johnson and Grossman (1977). Agarose gels of indicated concentration (T) were run at low (1.64 V/ cm) and high (6.58 V/cm) field strengths. Mobility increase values are the ratio M (6.58 V/cm) / M (1.64 V/cm). I, II, and III are supercoiled, nicked circular, and linear duplex forms of the indicated DNAs. MtDNA indicates mouse mitochondrial DNA.

ACKNOWLEDGEMENT

REFERENCES

Bailey, J. and Davidson, N. (1976). Methylmercury as a reversible denaturing agent for agarose gel electrophoresis, Analyt. Biochem. 70:75–85.

Bauer, W. R. (1978). Structure and reactions of closed duplex DNA, Annu. Rev. Biophys. Bioeng. 7:287–313.

Carle, G. F., Frank, M. and Olson, M. V. (1986). Electrophoretic separations of large DNA molecules by periodic inversion of the electric field, Science 232:65–68.

Chu, G., Vollrathe, D. and Davis, R. W. (1986). Separation of large DNA molecules by contour-clamped homogeneous electric field, Science 234:1582–1585.

Crick, F. H. C. (1976). Linking numbers and nucleosomes, Proc. Natl. Acad. Sci. U.S.A. 73:2639–2643.

Crick, F. H. C., Wang, J. C. and Bauer, W. R. (1979). Is DNA really a double helix? J. Mol. Biol. 129:449–461.

Depew, R. E. and Wang, J. C. (1975). Conformational fluctuations of DNA helix, Proc. Natl. Acad. Sci. U.S.A. 72:4275–4279.

Deutsch, J. M. (1988). Theoretical studies of DNA during gel electrophoresis, Science 240:922–924.

Deutsch, J. M. and Madden, T. L. (1989). Theoretical studies of DNA during gel electrophoresis, J. Chem. Phys. 90:2476–2485.

Fisher, M. P. and Dingman, C. W. (1971). Role of molecular conformation in determining the electrophoretic properties of polynucleotides in agarose-acrylamide composite gels, Biochemistry 10:1895–1899.

Flint, D. H. and Harrington, R. E. (1972). Gel electrophoresis of deoxyribonucleic acid, Biochemistry 11:4858–4864.

Fried, M. G. and Crothers, D. M. (1981). Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis, Nucleic Acids Res. 9:6505–6525.

Gariepy, C. E., Lomax, M. I. and Grossman, L. I. (1986). SPLINT: a cubic spline interpolation program for the analysis of fragment sizes in one-dimensional electrophoresis gels, Nucleic Acids Res. 14:575–581.

Helling, R. B., Goodman, H. M. and Boyer, H. W. (1974). Analysis of endonuclease R-EcoRI fragments of DNA from lambdoid bacteriophages and other viruses by agarose gel electrophoresis, J. Virol. 14:1235–1244.

Hsieh, C.-H. H. and Griffith, J. (1988). The terminus of SV40 DNA replication and transcription contains a sharp sequence-directed curve, Cell 52:535–544.

Hudspeth, M. E. S., Shumard, D. S., Bradford, C. J. R. and Grossman, L. I. (1983). Organization of Achlya mtDNA: a population with two orientations and a large inverted repeat containing the rRNA gene, Proc. Natl. Acad. Sci. U.S.A. 80:142–146.

Johnson, P. H. and Grossman, L. I. (1977). Electrophoresis of DNA in agarose gels: optimizing separations of conformational isomers of double- and single-stranded DNAs, Biochemistry 16:4217–4225.

Johnson, P. H., Miller, M. J., Wild, E., Kelly, S. V. and Grossman, L. I. (1979). Electrophoretic characterization of intracellular forms of bacteriophage DX174 DNA: identification of novel intermediates of altered superhelix density, J. Virol. 32:629–639.

Johnson, P. H., Miller, M. J. and Grossman, L. I. (1980). Electrophoresis of DNA in agarose gels: II. Effects of loading mass and electroendosmosis on electrophoretic mobilities, Analyt. Biochem. 102:159–162.

Keller, W. and Wendel, I. (1974). Stepwise relaxation of supercoiled SV40 DNA, Cold Spring Harbor Symp. Quant. Biol. 39:199–208.

Keller, W. (1975a). Characterization of purified DNA relaxing enzyme from human tissue culture cells, Proc. Natl. Acad. Sci. U.S.A. 72:2550–2554.

Keller, W. (1975b). Determination of the number of superhelical turns in simian virus 40 DNA by gel electrophoresis, Proc. Natl. Acad. Sci. U.S.A. 72:4876–4880.

Koepsel, R. R. and Khan, S. A. (1986). Static and initiator protein-enhanced bending of DNA at a replication origin, Science 233:1316–1318.

Klotz, L. C. and Zimm, B. H. (1972). Retardation times of deoxyribonucleic acid solutions. II. Improvements in apparatus and theory, Macromolecules 5:471–481.

Lai, E., Birren, B. W., Clark, S. M., Simon, M. and Hood, L. (1989). Pulsed field gel electrophoresis, Biotechniques 7:34–42.

Lee, C.-H., Mizusawa, H. and Kakefuda, T. (1981). Unwinding of double-stranded DNA helix by dehydration, Proc. Natl. Acad. Sci. U.S.A. 78:2838–2842.

Lerman, L. S., Fischer, S. G., Hurley, I., Silverstein, K. and Lumelsky, N. (1984). Sequence-determined DNA separations, Annu. Rev. Biophys. Bioeng. 13:399–423.

Linial, M. and Shlomai, J. (1987). Sequence-directed bent DNA helix is the specific binding site for Crithidia fasciculata nicking enzyme, Proc. Natl. Acad. Sci. U.S.A. 84:8205–8209.

McDonnell, M. W., Simon, M. N. and Studier, F. W. (1977). Analysis of restriction fragments of T7 DNA and determination of molecular weights by electrophoresis in neutral and alkaline gels, J. Mol. Biol. 110:119–146.

Maniatis, T., Jeffrey, A. and van de Sande, H. (1975). Chain length determination of small double- and single-stranded DNA molecules by polyacrylamide gel electrophoresis, Biochemistry 14:3787–3794.

Marini, J. C., Levene, S. D., Crothers, D. C. and Englund, P. T. (1982). Bent helical structure in kinetoplast DNA, Proc. Natl. Acad. Sci. U.S.A. 79:7664–7668.

Myers, R., Lumelsky, N., Lerman, L. S. and Maniatis, T. (1985). Detection of single base substitutions in total genomic DNA, Nature 313:495–498.

Myers, J. A., Sanchez, D., Elwell, L. P. and Falkow, S. (1976). Simple electrophoretic method for the identification and characterization of plasmid deoxyribonucleic acid, J. Bacteriol. 127:1529–1537.

Noll, W. W. and Collins, M. (1987). Detection of human DNA polymorphisms with a simplified denaturing gradient gel electrophoresis technique, Proc. Natl. Acad. Sci. U.S.A. 84:3339–3343.

Olivera, B. M., Baine, P. and Davidson, N. (1964). Electrophoresis of the nucleic acids, Biopolymers 2:245–257.

Pulleyblank, D. E., Shure, M., Tang, D., Vinograd, J. and Vosberg, H.-P. (1975). Action of nicking-closing enzyme on supercoiled and nonsupercoiled DNA: formation of a Boltzmann distri-

bution of topological isomers, Proc. Natl. Acad. Sci. U.S.A. 72:4280–4284.

Revzin, A. (1989). Gel electrophoresis assays for DNA-protein interactions, Biotechniques 7:346–355.

Richards, E. G. and Lecanidou, R. (1971). Quantitative aspects of the electrophoresis of RNA in polyacrylamide gels, Analyt. Biochem. 40:43–71.

Schwartz, D. C. and Cantor, C. R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis, Cell 37:67–75.

Schwartz, D. C. and Koval, M. (1989). Conformational dynamics of individual DNA molecules during gel electrophoresis, Nature 338:520–522.

Serwer, P. (1980). Electrophoresis of duples deoxyribonucleic acid in multiple concentration agarose gels: fractionation of molecules with molecular weights between 2 x 10^6 and 110 x 10^6, Biochemistry 19:3001–3004.

Serwer, P. (1983). Agarose gels: properties and use for electrophoresis, Electrophoresis 4:375–382.

Serwer, P. (1987). Gel electrophoresis with discontinuous rotation of the gel: an alternative to gel electrophoresis with changing direction of the electric field, Electrophoresis 8:301–304.

Serwer, P. and Allen, J. L. (1984). Conformation of double-stranded DNA during agarose gel electrophoresis: fractionation of linear and circular molecules with molecular weights between 3 x 10^6 and 26 x 10^6, Biochemistry 23:922–927.

Shure, M. and Vinograd, J. (1976). The number of superhelical turns in native virion SV40 DNA and minicol DNA determined by the band counting method, Cell 8:215–226.

Shure, M., Pulleyblank, D. E. and Vinograd, J. (1977). The problems of eukaryotic and prokaryotic DNA packaging and in vivo, Nucleic Acids Res. 4:1183–1205.

Smith, S. B., Aldridge, P. K. and Callis, J. B. (1989). Observation of individual DNA molecules undergoing gel electrophoresis, Science 243:203–206.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis, J. Mol. Biol. 98:503–517.

Southern, E. M. (1979). Gel electrophoresis of restriction fragments, Meth. Enzymol. 68:152–176.

Southern, E. M., Anand, R., Brown, W. R. A. and Fletcher, D. S. (1987). A model for the separation of large DNA molecules by crossed field gel electrophoresis, Nucleic Acids Res. 15:5925–5943.

Strauss, F. and Varshavsky, A. (1984). A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding into the nucleosome, Cell 37:889–901.

Vinograd, J., Lebowitz, J. and Watson, R. (1968). Early and late helix-coil transitions in closed circular DNA: the number of superhelical turns in polyoma DNA, J. Mol. Biol. 33:173–197.

Vollrath, D. and Davis, R. W. (1987). Resolution of DNA molecules greater than 5 megabases by contour-clamped homogeneous electric fields, Nucleic Acids Res. 15:7865–7876.

Vosberg, H.-P. (1985). DNA topoisomerases: enzymes that control DNA conformation, Curr. Topics Microbiol. Immunol. 114:19–102.

Wang, J. C. (1979). Helical repeat of DNA in solution, Proc. Natl. Acad. Sci. U.S.A. 76:200–203.

Wu, H.-M. and Crothers, D.M. (1984). The locus of sequence-directed and protein-induced DNA bending, Nature 308:509–513.

Zahn, K. and Blattner, F. (1985). Sequence-induced DNA curvature at the bacteriophage lambda origin of replication, Nature 317:451–453.

CLONING AND RECOMBINANT DNA TECHNOLOGIES FOR THE DEVELOPMENT OF HYBRIDIZATION PROBES

John S. Waye

Royal Canadian Mounted Police
Ottawa, Ontario, Canada

The advent of cloning and recombinant DNA technology (Cohen *et al.* 1973) has had widespread effects on basic and applied biological sciences. The term DNA cloning refers to the isolation a DNA sequence from virtually any source and its clonal propagation a heterologous host cell (bacterial, yeast, mammalian). To accomplish this, the DNA sequence must be in a form which can be faithfully replicated by the chosen cell host. Recombinant DNA technology can be used to insert a DNA sequence into an autonomously replicating extrachromosomal DNA vector, resulting in chimeric or recombinant DNA molecules which can be propagated in relatively large quantities.

Among the many impacts of molecular cloning and recombinant DNA technology is the ability to generate hybridization probes for the study of homologous sequences in the genome of origin or those of heterologous species. The focus of this brief review is recombinant DNA strategies for developing hybridization probes useful for the study of intra- and inter-species genetic variation. Particular emphasis will be placed on probes for the detection of restriction fragment length polymorphism (RFLP) of human DNA.

HYBRIDIZATION PROBES

A hybridization probe is simply a segment of single-stranded nucleic acid (DNA or RNA) that can be used in a hybridization reaction to identify complementary nucleic acid sequences. The specificity of the hybridization reaction is such that probes can be used to discriminate between complex mixtures of sequences (for example, eukaryotic genomes) and specifically identify the complementary sequence(s). To facilitate the detection of the probe:target hybrids, probes are generally labelled by *in vitro* incorporation of isotopic or non-isotopic conjugated nucleotides. The most common probe applications involve hybridization to DNA or RNA fragments that have been separated by gel electrophoresis and transferred to a membrane (Southern 1975; Thomas 1980). Other uses include hybridization to bacterial colonies (Grunstein and Hogness 1975) or phage plaques (Benton and Davis

1977) and hybridization *in situ* to intact chromosomes (Pardue and Gall 1969; Pinkel *et al.* 1986).

TYPES OF HUMAN DNA

From the perspective of probe development, perhaps the most important characteristic of any sequence is whether it is present in single or multiple copies per haploid genome. As much as 40% of the human genome consists of DNA sequences which are present in multiple copies (Britten and Davidson 1971; Britten and Kohne 1968), the remaining sequences comprising single-copy genes and non-coding inter-genic DNA.

By definition, single-copy DNA sequences define specific loci within a genome, a desirable feature for many probe applications. Single-copy sequences may be subdivided into those which are truly unique in the haploid genome and those that are related to other loci which derive from a common sequence. The latter class of single-copy sequences, best exemplified by the minisatellite (Jeffreys *et al.* 1985a,b) or variable number of tandem repeats (VNTR) (Nakamura *et al.* 1987) loci, are important since one can use probes that are specific for one locus to identify related located elsewhere in the genome.

In the relatively short time since the first RFLPs were described, strategies for RFLP identification (Skolnick and White 1982) and their use for construction of comprehensive genetic linkage maps (Botstein *et al.* 1980, White *et al.* 1985) have proved remarkably successful. To date, several thousand RFLPs have been described in the human genome (Pearson *et al.* 1987) and genetic linkage maps have been constructed for individual chromosomes (for example, Barker *et al.* 1987; Lathrop *et al.* 1988; Nakamura *et al.* 1988b) and for the genome as a whole (Donis-Keller *et al.* 1987). RFLP-based genetic linkage mapping has had its greatest impact on the study of inherited disorders, with linkage demonstrated between RFLPs and an ever increasing number of disease loci (reviewed by McKusick 1986). In addition, the availability of probes which detect highly polymorphic loci has provided powerful tools for human identity testing

(Balazs et al. 1989; Nakamura et al. 1987; Wainscoat et al. 1987; Wong et al. 1987).

Repetitive DNAs can be subdivided according to whether they are organized as extensive tandem arrays (satellite DNAs) or as solitary elements interspersed with single-copy sequences (reviewed by Singer 1982). Although there are several distinct types of satellite DNA in the human genome, these sequences are confined to a limited number of sites in the genome, principally the pericentromeric regions of chromosomes (Miklos and John 1979). Analogous to single-copy sequences, there are satellite sequences which define distinct domains yet are related to sequences found at more than one site. For example, the centromeric region of each human chromosome contains alpha satellite DNA, sequences which are all clearly derived from a common repeat yet have diverged into distinct chromosome-specific subsets (Willard 1985a; Willard and Waye 1987).

Notwithstanding the enormous value of single-copy RFLPs, probes which recognize highly repetitive satellite DNA may also serve as markers of intra- and inter-species genetic variation. It has long been known that satellite DNAs undergo rapid genetic turnover in a concerted manner (Dover 1982). Accordingly, these sequences are characterized by high degrees of intra-species homogeneity while exhibiting marked differences among even closely related species (for example, Bogenberger et al. 1987; Fanning et al. 1988; Waye and Willard 1989).

In addition to providing sensitive indicators of inter-species variability, many satellite DNAs are organized as chromosome-specific subsets and therefore may be used for the detection of individual chromosomes in situ or by Southern hybridization analysis. For example, there are several different types of satellite DNA which are specific for the Y chromosome (Willard 1985b; Wolfe et al. 1985) and can serve as male-specific markers (for example, Gill 1987; Lau et al. 1984; Tyler et al. 1986). Lastly, there are satellite DNA loci that are polymorphic and may be useful for gene mapping (Willard et al. 1986).

Interspersed repetitive DNAs are found throughout most of genome, the exception being that they are substantially under-represented within the aforementioned domains of satellite DNA (Moyzis et al. 1989). Due to a relatively high degree of sequence identity shared among members of individual interspersed repetitive DNA families (Deininger et al. 1981), it is technically difficult to use these sequences as genetic markers. However, these types of sequences exhibit degrees of species-specific homogeneity (Daniels et al. 1983) and may be used to differentiate among mammalian species (Tyler et al. 1986).

GENERAL STRATEGIES FOR MOLECULAR CLONING OF GENOMIC DNA SEQUENCES

Most cloning strategies represent variations of a common approach, beginning with the construction of recombinant DNA molecules consisting of the DNA sequence of interest and a suitable cloning vector. A cloning vector is simply a DNA molecule that is capable of autonomous replication in a given host cell (for example, bacteria, yeast, cultured mammalian or plant cells). The vector DNA contains sequences which control its replication (origin of replication) as well as regions within which DNA sequences may be inserted (cloning sites). In addition, some vectors contain selectable markers such as genes which confer antibiotic resistance.

A most simple version of the recombinant DNA process relies on restriction endonucleases to generate compatible vector and insert termini (for example, vector and insert termini generated with the same restriction endonuclease, isoschizomers, or enzyme combinations which generate blunt ends or complementary 3' or 5' overhangs). The fragments are ligated in vitro using T4 ligase, an enzyme that catalyzes the formation of phosphodiester bonds between 3' hydroxyl and 5' phosphate termini of DNA. Following ligation, the recombinant molecule (vector plus inserted DNA) is introduced into a suitable host cell. At this point, the recipient cells are said to be transformed and the recombinant molecule may be clonally replicated. Although a detailed description of the various cloning vectors and ligation strategies is beyond the scope of this review (see Ausubel et al. 1987; Berger and Kimmel 1987; Davis et al. 1986; Dillon et al. 1985; Glover 1987; Maniatis et al. 1982), the salient features of the most widely employed vector-host systems will be surveyed.

The most versatile vectors are those designed for use in the bacterium Escherichia coli (E. coli). These include plasmids, bacteriophage lambda, and cosmids. Plasmids are small circular DNA molecules which replicate as extrachromosomal episomes and may accommodate several kilobase pairs (Kb) of foreign DNA (reviewed by Bolivar and Backman 1979). Recombinant plasmids can be introduced into E. coli using simple chemical pretreatments to render the bacterial cells competent for DNA uptake (Mandel and Higa 1970; Hanahan

1983). After DNA uptake, the plasmids replicate autonomously and can be harvested by any number of lytic treatments (for example, Birnboim and Doly 1978).

Bacteriophage lambda is a temperate phage (bacterial virus) whose genome can accommodate 10–20 Kb of foreign DNA (Leder et al. 1978; Tremeier et al. 1976). Recombinant phage genomes can be packaged into phage particles in vitro and used to infect E. coli strains. Once infected, the cell makes many copies of the phage genome which, in turn, are packaged into phage particles and released upon cell lysis.

Cosmids are cloning vectors which have features of both plasmids and bacteriophage lambda and can accommodate up to 40 Kb of foreign DNA (Collins and Hohn 1978). They contain a plasmid origin of replication, selectable marker, and cloning sites. They also contain bacteriophage sequences (Cos sites) which allow the molecules to be packaged into phage particles in vitro. The cosmid-containing phage particles can then be used for efficient infection of E. coli strains. Once introduced into the bacterial host, the cosmid DNA replicates much the same as a large selectable plasmid.

Lambda and cosmid vectors have two distinct advantages over plasmids for constructing extensive collections, or libraries, of recombinant molecules. Lambda and cosmid vectors can accommodate larger DNA inserts than plasmids and therefore a smaller number of recombinant clones is needed to adequately represent all of the DNA sequences of a given genome. Although there are no physical or packaging constraints on the size of plasmids, reduced transformation and replication efficiencies are associated with inordinately large plasmid constructs. Secondly, the efficiency with which phage particles can be introduced into E. coli is substantially greater than that for plasmids. Phage particles containing recombinant lambda or cosmid DNA infect E. coli with efficiencies approaching 100%. By comparison, plasmid transformation is less efficient, with routine transformation frequencies of 10^8 transformants per μg plasmid or <1 in 1000 plasmid molecules being successfully introduced into the bacterial host.

The advantages just discussed (large insert sizes and efficient infection) make bacteriophage lambda and cosmids the preferred vectors for cloning DNA segments from large genomes. Random genomic fragments can be cloned into the vectors and the resulting genomic library can then be accessed for individual DNA segments (for example,

Blattner et al. 1978; Grosveld et al. 1980; Maniatis et al. 1978). Plasmids, on the other hand, are more commonly used as vectors for subcloning the desired portions of DNA segments derived from individual lambda or cosmid clones.

CLONING DNA PROBES FOR DETECTION OF SINGLE-COPY RFLPS

Once a comprehensive genomic DNA library has been constructed (that is, one containing enough fragments to adequately represent the genome) and introduced into the appropriate host cells, one must somehow identify the recombinant sequences which correspond to polymorphic single-copy loci. One approach is to simply select clones (bacterial colonies or phage plaques) at random and use the individual recombinant clones as hybridization probes against genomic DNAs digested with different restriction endonucleases. For efficient RFLP screening, genomic DNA from a limited number of unrelated individuals is analyzed with a number of restriction endonucleases, concentrating on those enzymes whose recognition sites are most likely to be polymorphic (Barker and White 1982; Cooper and Schmidtke 1985). This strategy, albeit labor intensive, has been used to isolate several hundred RFLP probes (Bowden et al. 1989; Schumm et al. 1988). A drawback to this approach is that many of the RFLPs will be simple bi-allelic polymorphisms with limited discrimination potential.

A second approach to identifying polymorphic loci is to use a DNA probe to screen the recombinant library for clones bearing homology to the probe sequence (Benton and Davis 1977; Grunstein and Hogness 1975; Maniatis et al. 1978). This approach has been used to identify clones bearing homology to minisatellite or VNTR core sequences. The VNTRs are a family of interspersed loci which are based on short tandem arrays of a common core sequence (Nakamura et al. 1987). There are several thousand of these loci located throughout the human genome and they can exhibit high degrees of polymorphism due to variable numbers of tandem repeats in each array. The fact that these loci are based on a common core sequence has been exploited in schemes designed to isolate VNTR probes. Nakamura and co-workers have isolated clones which identify VNTR loci by screening cosmid libraries with synthetic oligonucleotide probes corresponding to different VNTR core sequences (Nakamura et al. 1987, 1988a). This approach could also be used to identify loci containing arrays of the dinucleotide repeat (TG)n, re-

cently shown to represent another type of poly-morphic variation (Litt and Luty 1989; Weber and May 1989).

Wong *et al.* (1986, 1987) have used a directed approach for the cloning of particular VNTR sequences identified as hypervariable DNA fragment lengths in multilocus DNA fingerprint patterns. They constructed a lambda library from a purified preparation of genomic fragments whose lengths coincide with the polymorphic fragment lengths. The library was screened with a minisatellite probes and locus-specific probes were subsequently identified. This approach is attractive since it is directed against the isolation of fragments which *a priori* are known to be highly polymorphic.

PROBE MODIFICATIONS—SUBCLONING STRATEGY FOR REMOVAL OF REPETITIVE DNAS FROM CLONED PROBES

As a consequence of the large insert sizes permitted in lambda and cosmid vectors, it is not uncommon that individual clones contain one or more interspersed repetitive elements as part of the genomic DNA insert. This is undesirable since these sequences are a component of the DNA probe and will hybridize to homologous interspersed elements throughout the genome. This has the effect of masking the single-copy hybridization pattern identified by the unique sequence components of the probe. To overcome this problem, the probe sequences homologous to interspersed repetitive elements can be preferentially quenched by hybridization in the presence of excess unlabelled genomic DNA (Litt and White 1985). While this measure permits detection of the single-copy hybridization pattern, it may also result in reduced sensitivity due to limited quenching of the single-copy probe sequences. As such, this is generally employed only to evaluate the degree of polymorphism detected by individual lambda or cosmid clones. Once it is established that a particular clone detects an RFLP, efforts are usually made to construct plasmid subclones of the single-copy sequences.

Single-copy sequences within large genomic inserts are most easily identified using restriction endonucleases to subdivide the insert DNA. The restriction fragments are then analyzed by Southern hybridization using total human genomic DNA or cloned interspersed repetitive DNAs as hybridization probes. The insert fragments which do not contain repetitive elements (that is, those containing only single-copy DNA) will not exhibit any hybridization and can be isolated and subcloned in appropriately digested plasmid vectors.
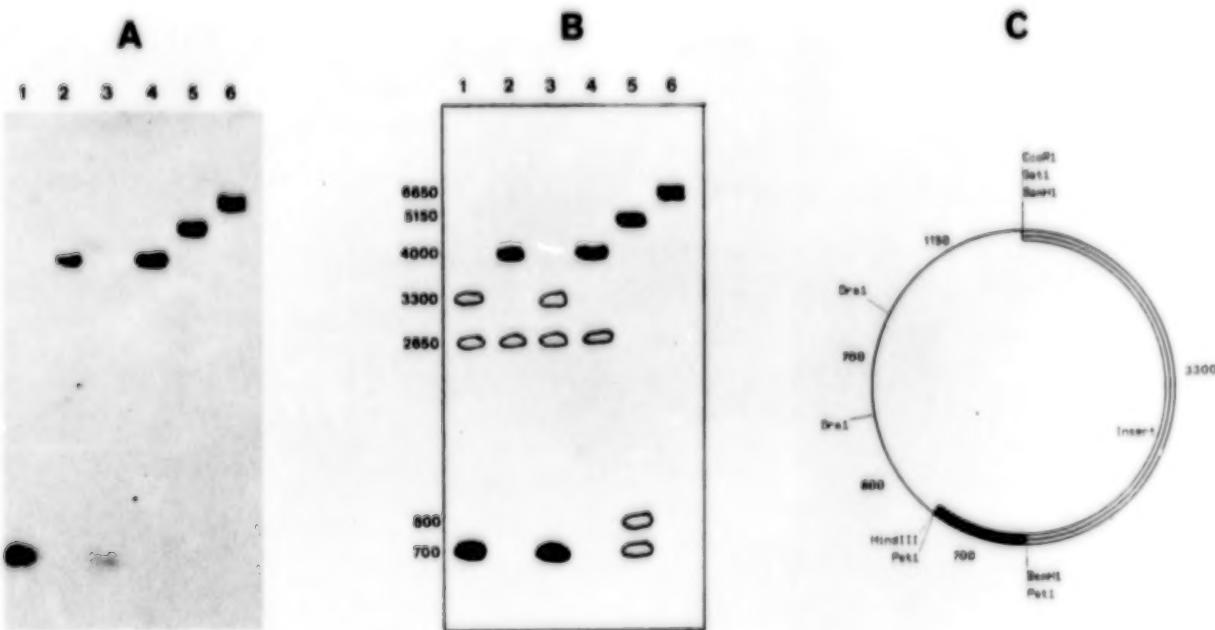
Figure 1: Probe modification to eliminate repetitive elements. (A) Plasmid pEFD64.2 was digested with several different restriction endonucleases and analyzed by Southern hybridization with labelled total human genomic DNA. Lane 1 = *Pst*I-*Eco*RI; lane 2 = *Eco*RI-*Hind*III; lane 3 = *Bam*HI-*Hind*III; lane 4 = *Sst*-*Hind*III; lane 5 = *Dra*I-*Hind*III; lane 6 = *Eco*RI. (B) Schematic of the pattern of fragments generated by the restriction analysis shown in (A). The open bands are those which failed to hybridize with the total human DNA probe and the closed bands are those which did hybridize. The sizes of the bands are indicated in base pairs. (C) Restriction map of pEFD64.2. The genomic insert is indicated and the portion which contains the repetitive element is highlighted.

Figure 1 demonstrates this approach for modification of a clone that contains VNTR sequences as well as an interspersed repetitive element. The original clone (pEFD64.2, Nakamura *et al.* 1988b) contains a ~4.0 Kb insert, consisting of ~3.3 Kb of single-copy DNA and a ~0.7 Kb fragment that harbors a repetitive DNA element. As a hybridization probe, the intact 4.0 Kb insert hybridized extensively to repetitive sequences in human genomic DNA, effectively masking single-copy hybridization signals in the low molecular weight range of the blot (Figure 2A). However, when the 3.3 Kb single-copy portion of the probe was subcloned and used as probe, there was no lane background and the single-copy hybridization pattern was no longer obscured (Figure 2B).

CLONING DNA PROBES FOR DETECTION OF REPETITIVE DNA SEQUENCES

As with single-copy DNAs, repetitive DNAs can be isolated from genomic libraries using cross-reactive probes to identify homologous recombinant clones. This approach, however, suffers from the fact that satellite DNAs are often under-repre-
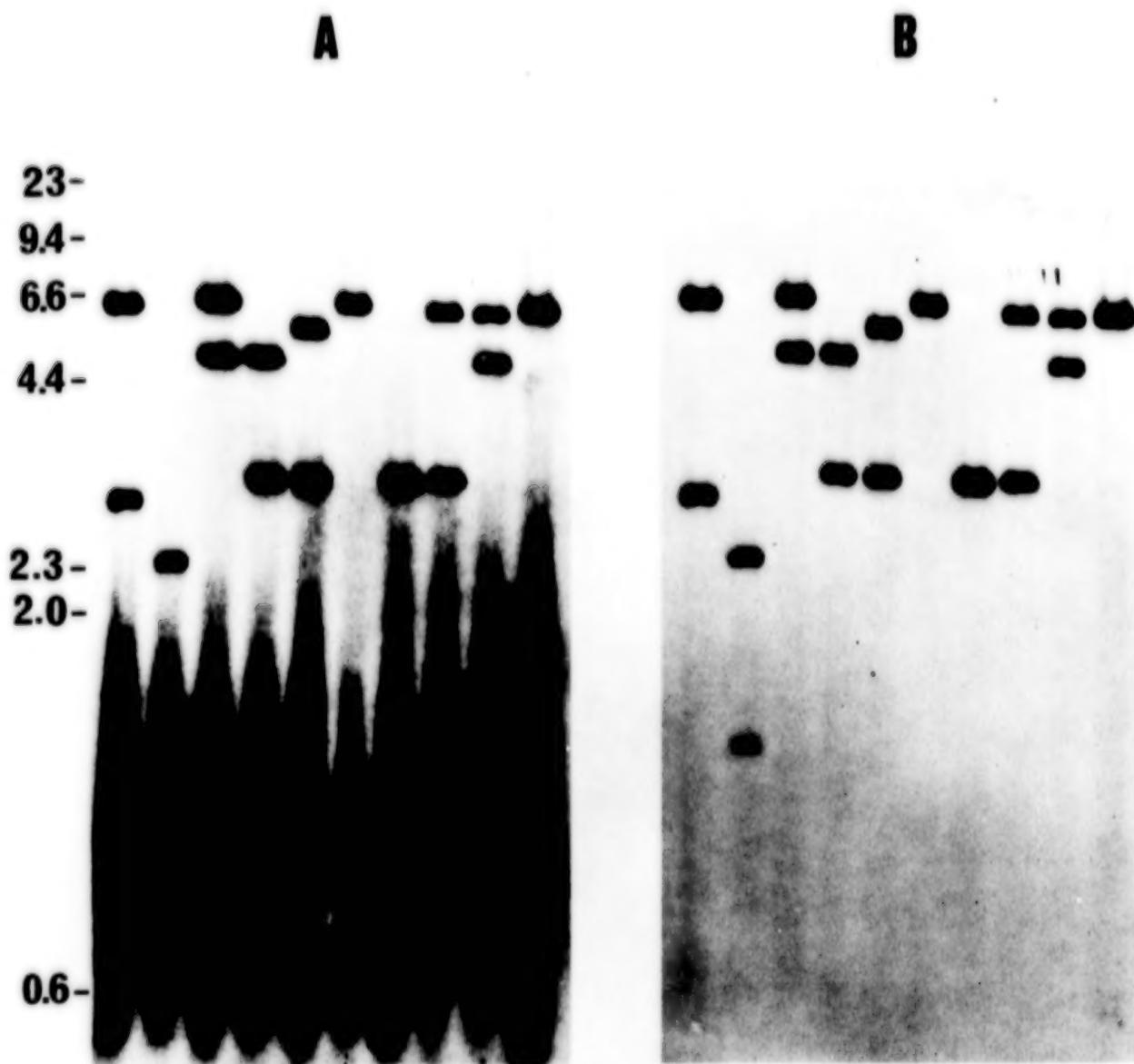
Figure 2: Hybridization of with a probes containing a repetitive elements and a single-copy derivative. Genomic DNA of ten unrelated individuals analyzed with HinfI and probed with the entire 4.0 Kb *EcoR*I-*Hind*III insert (panel A) or the 3.3 *BamH*I insert (panel B) of pEFD64.2 (see Figure 2). Sizes are indicated in Kb.

sented in genomic libraries due to the lack of the particular restriction site used for library construction. For example, if a tandem repeat contains no *Sau3*A sites, there will be a paucity of *Sau3*A sites throughout the array and any library based on *Sau3*A inserts will not contain that particular satellite sequence.

Other strategies for the isolation of satellite DNA probes take advantage of the fact that these sequences consist of highly repetitive tandem sequences. Repetitive DNAs (satellite or interspersed) reassociate faster than low- or single-copy DNAs in classical denaturation-annealing experiments (Britten and Kohn 1968). This property of repetitive DNA has been exploited by Moyzis *et al.* (1987) to create a human repetitive DNA library from which several different types of repetitive DNAs were subsequently isolated.

Satellite DNAs may also be isolated based on differences in base composition relative to the bulk of genomic DNA sequences. If a satellite repeat has a base composition which significantly differs from genomic DNA as a whole, it will separate as a distinct satellite band following isopycnic density centrifugation (for example, Kit 1961; Schildkraut *et al.* 1962). Satellite DNAs isolated in this manner may be used as uncloned probes or can be cloned into a suitable vector following the appropriate treatments to ensure compatibility of the vector and insert termini.

A more direct approach for isolation of satellite DNA sequences takes advantage of the observation that a tandem array may be reduced to its fundamental unit of tandem reiteration by restriction endonucleases which cleave within the repeating unit. For example, the major satellite DNA of the Y chromosome consists of several thousand copies of a 3.4 Kb tandem repeat unit within which lies a unique *Hae*III site. Accordingly, cleavage with *Hae*III reduces the array to several thousand 3.4 Kb *Hae*III fragments which can be detected as a visible band in ethidium bromide-stained male genomic DNA (Cooke 1976).

Once detected, a satellite band can be isolated from the electrophoretic gel and cloned directly into plasmid vectors. Since the isolated size fraction is highly enriched for the particular repeat unit, it may be used as an uncloned probe to screen the plasmid library for the corresponding recombinant molecules. In this way, it is possible to clone satellite sequences with no previous knowledge as to the nature of the sequence. Figure 3 illustrates this approach for cloning species-specific satellite DNA probes. From gorilla genomic DNA, a 0.34

Kb *Eco*RI satellite band was isolated and used to construct a small plasmid library. Cloned copies of the 0.34 Kb repeats were identified on the basis of hybridization using the uncloned 0.34 Kb *Eco*RI fragment as a probe (Waye *et al.* 1989). Both the uncloned and cloned 0.34 *Eco*RI repeats were shown to be gorilla-specific under the appropriate hybridization conditions (Figure 3). As a corollary, the same experiments with human genomic DNA yielded uncloned and cloned probes which are human-specific under high stringency. Since satellite DNAs are a ubiquitous property of eukaryotic genomes (Beridze 1986), it should be possible to apply this approach to virtually any species.

OVERVIEW—A FORENSIC PERSPECTIVE

The availability of cloned DNA fragments for use as hybridization probes promises to revolutionize forensic identity testing. Without any *a priori* knowledge of the origin or history of a forensic sample, it is possible to use highly repetitive human-specific probes to identify and quantify human genomic DNA (Waye *et al.* 1989). In addition, repetitive probes specific for the Y chromosome can be used to ascertain the sex of the DNA donor. Following these initial profiling tests, the DNA can be analyzed with a succession of probes, each of which recognize independent hypervariable loci and collectively provide effective individualization.

To date, recombinant DNA technology has provided cloned probes corresponding to numerous repetitive and single-copy hypervariable loci in the human genome (Pearson *et al.* 1987). As the number of cloned RFLP probes continues to increase at a remarkable rate, one might wonder whether the discovery of an additional 100, or even 1000 hypervariable RFLP probes change the direction of forensic DNA typing? In the short term, the availability of new probes will allow for RFLP detection of genetically independent hypervariable loci with a single restriction endonuclease, as well identify VNTR loci suitable for enzymatic amplification (Jeffreys *et al.* 1988). Arguably, these objectives are not far from being realized since there are numerous VNTR loci suitable for RFLP detection and enzymatic amplification.

A note of future promise for recombinant DNA technology may be the application of similar approaches for the identity testing in non-human species. In particular, this methodology may be used to develop identity tests for domestic, wildlife, and endangered species.
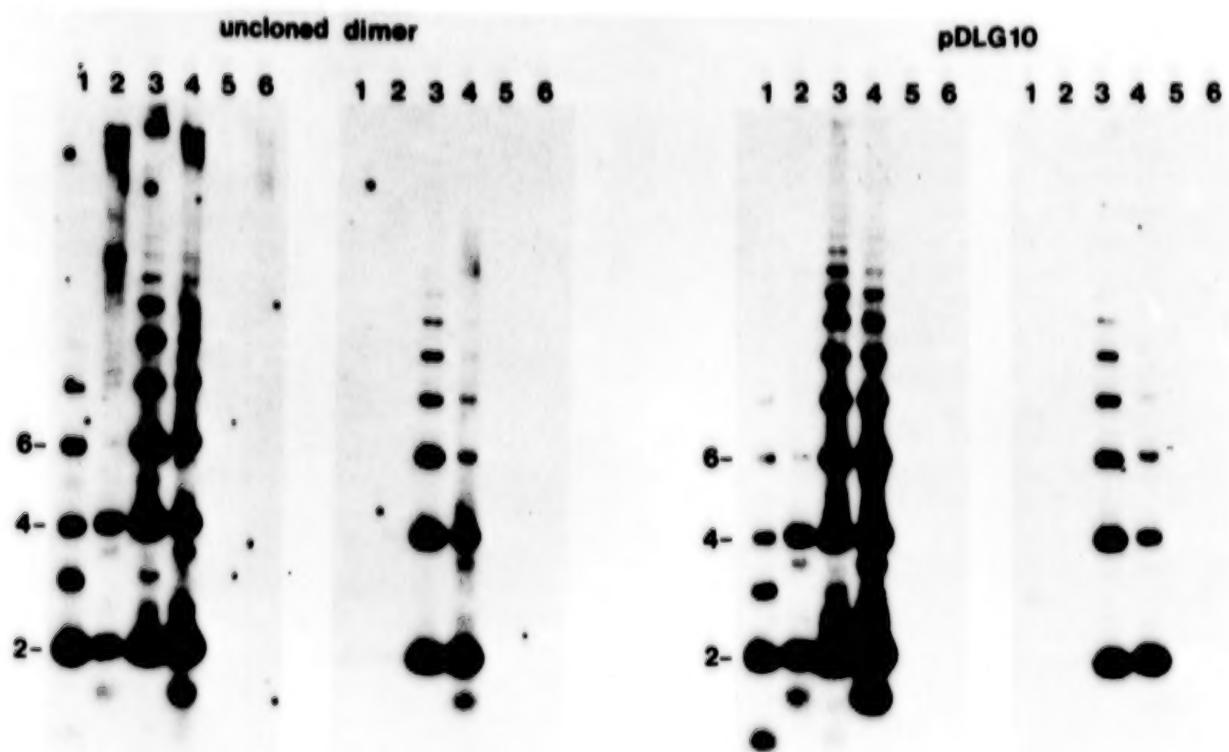
Figure 3: Isolation of species-specific satellite DNA probes. A visible 0.34 Kb satellite fragment was isolated from EcoRI-digested gorilla genomic DNA separated on a preparative 1.5% agarose gel. From this fragment, a plasmid library was constructed and the uncloned probe was used to identify recombinants containing homologous sequences (for example, pDLG10). The uncloned and cloned probes were hybridized at low stringency (left panels) and high stringency (right panels) to digests of human (lanes 1 and 2), gorilla (lanes 3 and 4), and orangutan genomic DNAs (lanes 5 and 6). Band sizes are indicated in multiples of the 0.34 Kb tandem repeat unit.

REFERENCES

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1987). Current Protocols in Molecular Biology. John Wiley and Sons, NY. ·

Balazs, I., Baird, M., Clyne, M. and Meade, E. (1989). Human population genetic studies of five hypervariable DNA loci, Am. J. Hum. Genet. 44:182–190.

Barker, D., Green, P., Knowlton, P., Schumm, J., Lander, E., Oliphant, A., Willard, H., Akots, G., Brown, V., Gravius, T., Helms, C., Nelson, C., Parker, C., Rediker, K., Rising, M., Watt, D., Weiffenbch, B. and Donis-Keller, H. (1987). A genetic map of 63 markers on human chromosome 7, Proc. Natl. Acad. Sci. USA 84:8006–8010.

Barker, D. and White, R. (1982). More base pair change polymorphisms at sites containing CpG. Cytogenet, Cell Genet. 32:253.

Benton, W. D. and Davis, R. W. (1977). Screening gt recombinant clones by hybridization to single plaques in situ, Science 196:180–182.

Berger, S. L. and Kimmel, A. R. (1987). Guide to molecular cloning techniques. In: Methods in Enzymology Vol 152. Academic Press, San Diego.

Beridze, T. (1986). Satellite DNA. Springer-Verlag, NY.

Birnboim, H. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA, Nucleic Acids Res. 7:1513–1523.

Blattner, F. R., Blechl, A. E., Denneston-Thompson, K., Faber, H. E., Richards, J. E., Slightom, J. L., Tucker, R. W. and Smithies, O. (1978). Cloning of human fetal gamma-globin and mouse alpha type globin DNA preparation and screening of shotgun collections, Science 202:1279–1284.

Bogenberger, J. M., Neitel, H. and Fittler, F. (1987). A highly repetitive component common to all Cervidae: its organization and chromosomal distribution during evolution, Chromosoma 95:154–161.

Bolivar, F. and Backman, K. (1979). Plasmids of *Escherichia coli* as cloning vectors, Methods Enzymol. 68:245–267.

Botstein, D., White, R., Skolnick, M. and Davis, R. W. (1980). Construction of genetic linkage maps using restriction fragment length polymorphism, Am. J. Hum. Genet. 32:314–331.

Bowden, D. W., Muller-Kahle, H., Gravius, T. C., Helms, C., Watt-Morgan, D., Green, P. and Donis-Keller, H. (1989). Identification and characterization of 23 RFLP loci by screening random cosmid genomic clones, Am. J. Hum. Genet. 44:671–678.

Britten, R. J. and Davidson, E. A. (1971). Repetitive and non-repetitive sequences and a speculation on the origins of evolutionary novelty, Quart. Rev. Biol. 46:111–133.

Britten, R. J. and Kohne, D. E. (1968). Repeated sequences in DNA, Science 161:529–540.

Cohen, S. N., Chang, A. C. Y., Boyer, H. W. and Heeling, R. B. (1973). Construction of functional bacterial plasmids in vitro, Proc. Natl. Acad. Sci. USA 70:3240–3244.

Collins, J. and Hohn, B. (1978). Cosmids: a type of plasmid gene-cloning vector that is packagable in vitro in bacteriophage lambda heads, Proc. Natl. Acad. Sci. USA 75:4242–4246.

Cooke, H. J. (1976). Repeated sequence specific to human males, Nature 262:182–186.

Cooper, D. and Schmidtke, J. (1985). DNA RFLPs and heterozygosity in the human genome, Hum. Genet. 69:201–205.

Daniels, G. R., Fox, G. M., Loewensteiner, D., Schmid, C. W. and Deininger, P. L. (1983). Species-specific homogeneity of primate Alu family of repeated DNA sequences, Nucleic Acids Res. 11:7579–7593.

Davis, L. G., Dibner, M. D. and Batty, J. F. (1986). Basic Methods in Molecular Biology. Elsevier, NY.

Deininger, P. L., Jolly, D. J., Rubin, C. M., Friedmann, T. and Schmid, C. W. (1981). Base sequence studies of 300 nucleotide renatured repeated human DNA clones, J. Mol. Biol. 151:17–33.

Dillon, J.-A. R., Nasim, A. and Nestman, E. R. (1985). Recombinant DNA Methodology. John Wiley and Sons, NY.

Donis-Keller, H. and 32 co-authors. (1987). A genetic linkage map of the human genome, Cell 51:319–337.

Dover, G. (1982). Molecular drive: a cohesive mode of species evolution, Nature 299:111–116.

Fanning, T. G., Modi, W. S., Wayne, R. K. and O'Brien, S. J. (1988). Evolution of heterochro-matin-associated satellite DNA loci in felids and canids (Carnivora), Cytogenet Cell. Genet. 48:214–219.

Gill, P. (1987). A new method for sex determination of the donor of forensic samples using a recombinant DNA probe, Electrophoresis 8:35–38.

Glover, D. M. (1987). DNA Cloning, Vol III. A Practical Approach. IRL Press, Oxford.

Grosveld, F. G., Dahl, H.-H. M., deBoer, E. and Flavell, R. A. (1980). Isolation of beta-globin-related genes from a human cosmid library, Gene 13:227–237.

Grunstein, M. and Hogness, D. (1975). Colony hybridization: a method for the isolation cloned DNAs that contain a specific gene, Proc. Natl. Acad. Sci. USA 72:3961–3965.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids, J. Mol. Biol. 166:557–580.

Jeffreys, A. J., Wilson, V., Neumann, R. and Keyte, J. (1988). Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells, Nucleic Acids Res. 16:10953–10971.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985a). Individual-specific "fingerprints" of human DNA, Nature 316:76–79.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985b). Hypervariable "minisatellite" regions in human DNA, Nature 314:67–73.

Kit, S. (1961). Equilibrium sedimentation in density gradient of DNA preparations from animal tissues, J. Mol. Biol. 3:711–716.

Lathrop, G. M., Nakamura, Y., Cartwright, P., O'Connell, P., Leppert, M., Jones, C., Tateishi, H., Bragg, T., Lalouel, J.-M. and White, R. (1988). A primary genetic map of markers for human chromosome 10, Genomics 2:157–164.

Lau, Y.-F., Dozy, A. M., Huang, J. C. and Kan, Y. W. (1984). A rapid screening test for antenatal sex determination, Lancet I:14–16.

Leder, P., Tremeier, D. and Enquist, L. (1977). EK2 derivatives of bacteriophage lambda useful for cloning of DNA from higher organisms: the gtWES system, Science 196:175–177.

Litt, M. and Luty, J. A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene, Am. J. Hum. Genet. 44:397–401.

Litt, M. and White, R. (1985). A highly polymorphic locus in human DNA revealed by

cosmid-derived probes, Proc. Natl. Acad. Sci. USA 82:6206–6210.

Maniatis, T., Fritsch, E. F. and Sambrook. J. (1982). Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Maniatis, T., Hardison, R. C., Lacy, E., Lamer, J., O'Connell, C., Quan, D., Sim, G. U. and Efstratiadis, A. (1978). The isolation of structural genes from libraries of eukaryotic DNA, Cell 15:687–701.

Mandel, M. and Higa, A. (1970). Calcium-dependent bacteriophage DNA infection, J. Mol. Biol. 53:154.

McKusick, V. A. (1986). Mendelian inheritance in man: catalogs of autosomal dominant, autosomal recessive, and X-linked phenotypes, 7th edition. Johns Hopkins University Press, Baltimore.

Miklos, G. L. G. and John, B. (1979). Heterochromatin and satellite DNA in man: properties and prospects, Am. J. Hum. Genet. 31:264–280.

Moyzis, R. K., Albright, K., Bartholdi, M., Cram, L., Deaven, L., Hindebrand, C., Joste, N., Longmire, J. L., Meyne, J. and Schwarzachen-Robinson, T. (1987). Human chromosome-specific repetitive DNA sequences, Chromosoma 95:375–386.

Moyzis, R. K., Torney, D. C., Meyne, J., Buckingham, J. M., Wu, J.-R., Burks, C., Sirotkin, K. M. and Goad, W. B. (1989). The distribution of interspersed repetitive DNA sequences in the human genome, Genomics 4:273–289.

Nakamura, Y., Carlson, M., Krapcho, K., Kanamori, M. and White, R. (1988a). A new approach for isolation of VNTR markers, Am. J. Hum. Genet. 43:854–859.

Nakamura, Y., Fujimoto, E., O'Connell, P., Leppert, M., Lathrop, G. M., Lalouel, J.-M. and White, R. (1988b). Isolation and characterization of a polymorphic DNA sequence pEFD64.2 on chromosome 3 [D3S46], Nucleic Acids Res. 16:9354.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin,. E. and White, R. (1987). Variable number of tandem repeats (VNTR) markers for human gene mapping, Science 235:1616–1622.

Pardue, M. L. and Gall, J. G. (1969). Formation and detection of RNA-DNA hybrid molecules in cytological preparations, Proc. Natl. Acad. Sci. USA 63:378–383.

Pearson, P. L., Kidd, K. K. and Willard, H. F. (1987). Human gene mapping by recombinant DNA techniques, Cytogenet. Cell Genet. 46:390–566.

Pinkel, D., Straume, T. and Gray, J. W. (1986). Cytogenetic analysis using quantitative, high-sensitivity fluorescence hybridization, Proc. Natl. Acad Sci. USA 83:2934–2938.

Schildkraut, C. L., Marmur, J. and Doty, P. (1962). Determination of the base composition of deoxyribonucleic acid from its buoyant density in CsCl, J. Mol. Biol. 4:430–443.

Schumm, J., Knowlton, R., Braman, J., Barker, D., Botstein, D., Akots, G., Brown, V., Gravius, T., Helms, C., Hsaio, K., Rediker, K., Thurston, J. and Donis-Keller, H. (1988). Identification of more than 500 restriction fragment length polymorphisms by screening random genomic clones, Am. J. Hum. Genet. 42:143–159.

Singer, M. F. (1982). Highly repeated sequences in mammalian genomes, Int. Rev. Cytol. 76:67–112.

Skolnick, M. and White, R. (1982). Strategies for detecting and characterizing restriction fragment length polymorphisms (RFLPs), Cytogenet. Cell Genet. 32:58–67.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis, J. Mol. Biol. 98:503–517.

Thomas, P. S. (1980). Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose, Proc. Natl. Acad. Sci. USA 77:5201–5205.

Tremeier, D., Enquist, L. and Leder, P. (1976). An improved derivative of bacteriophage lambda EK2 vector useful for cloning of recombinant molecules: gtWES, Nature 263:526–527.

Tyler, M. G., Kirby, L. T., Wood, S., Vernon, S. and Ferris, J. A. J. (1986). Human blood stain identification and sex determination in dried blood stains using recombinant DNA techniques, Forensic Sci. Int. 31:267–272.

Wainscoat, J. S., Pilkington, S., Petro, T. E. A., Bell, J. I. and Higgs, D. R. (1987). Allele-specific DNA identity patterns, Hum. Genet. 75:384–387.

Waye, J. S., Presley, L., Budowle, B., Shutler, G. G. and Fourney, R. M. (1989). A simple and sensitive method for quantifying human genomic DNA in forensic specimen extracts, Biotechniques (in press).

Waye, J. S. and Willard, H. F. (1989). Concerted evolution of alpha satellite DNA: evidence for species-specificity and a general lack of se-

quence conservation among alphoid sequences of higher primates, Chromosoma (in press).

Weber, J. L. and May, P. E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction, Am. J. Hum. Genet. 44:388–396.

Willard, H. F. (1985a). Chromosome-specific organization of human alpha satellite DNA, Am. J. Hum. Genet. 37:524–532.

Willard, H. F. (1985b). Molecular organization of repeated sequences on the human Y chromosome. In: The Y Chromosome, Part A: Basic Characteristics of the Y Chromosome. Alan R. Liss, NY.

Willard, H. F. and Waye, J. S. (1987). Hierarchical order in chromosome-specific human alpha satellite DNA, Trends Genet. 3:192–198.

Willard, H. F., Waye, J.S., Skolnick, M., Schwartz, C. E., Powers, V. E. and England, S. B. (1986). Detection of restriction fragment length polymorphisms at the centromeres of human chromosomes using chromosome-specific alpha satellite DNA: implications for the development of centromere-based genetic linkage maps, Proc. Natl. Acad. Sci. USA 83:5611–5615.

White, R., Leppert, M., Bishop, D. T., Barker, D., Berkowitz, J., Brown, C., Callahan, P., Holm, T. and Jerominski, L. (1985). Construction of linkage maps with DNA markers for human chromosomes, Nature 313:101–105.

Wolfe, J., Darling, S. M., Erickson, R. P., Craig, I. W., Buckle, V. J., Rigby, P. W. J., Willard, H. F. and Goodfellow, P. N. (1985). Isolation and characterization of an alphoid centromeric repeat family from the human Y chromosome, J. Mol. Biol. 182:477–485.

Wong, Z., Wilson, V., Jeffreys, A. J. and Thein, S. L. (1986). Cloning a selected fragment from a human DNA "fingerprint": isolation of an extremely polymorphic minisatellite, Nucleic Acids Res. 14:4605–4616.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. J. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA, Ann. Hum. Genet. 51:269–288.

A PROTOCOL FOR RFLP ANALYSIS OF FORENSIC BIOSPECIMENS

Bruce Budowle

FBI Laboratory

Quantico, Virginia

Molecular biology techniques that reveal genetic polymorphisms at the level of the DNA are revolutionizing forensic science. One technology that already figures prominently for forensic analysis of biospecimens is that which detects restriction fragment length polymorphisms (RFLPs). Usually a RFLP gives rise to only two alleles at a locus due to the creation or elimination of a restriction site. This type of polymorphism does not provide a high degree of discrimination for characterization of evidentiary materials. However, with the advent of probes that recognize variable number of tandem repeats (VNTRs) (Wyman and White 1980; Nakamura *et al.* 1987a; Wong *et al.* 1987) highly polymorphic systems for DNA analysis are now available (Nakamura *et al.* 1987a,b, 1988a,b; Wong *et al.* 1987; Odelberg *et al.* 1989; Balazs *et al.* 1989; Milner *et al.* 1989; Wainscoat *et al.* 1987; Donnis-Keller *et al.* 1987). Some VNTR-RFLP systems are so polymorphic that three to five loci can be analyzed and provide unique identity (excluding monozygotic twins). In fact, presently, analysis of VNTR loci are the best effort a forensic scientist can make to eliminate an individual who has been falsely associated with a particular sample. A further advantage of this technology is it has a sensitivity of detection sufficient enough to be applicable to the limited quantities of DNA encountered in forensic samples (Wong *et al.* 1987; Budowle and Baechtel 1989).

While numerous procedures exist for RFLP analysis (Denhardt 1966; Southern 1975; Wahl *et al.* 1979; Maniatis *et al.* 1982; Gatti *et al.* 1984; Gillespie and Spiegelman 1985), all are modifications of a basic theme. The following outline displays the major steps required for a RFLP analysis.

1) Extract and purify DNA
2) Enzymatically digest DNA
3) Fractionate DNA fragments via electrophoresis
4) Transfer DNA from an electrophoretic gel to a membrane support via Southern Blotting (Southern 1975).
5) Probe membrane-bound DNA
6) Assay to detect probe:target DNA duplexes
7) Interpret patterns

Many protocols for RFLP analysis employ unnecessary steps and excessive reagents that have little impact on the end result or could even have a detrimental effect (Budowle and Baechtel 1989). For the information that can be obtained from a VNTR-RFLP analysis of a forensic sample to be derived routinely, a streamlined and efficacious method was developed that is suited to the demands of the forensic serology laboratory.

METHODOLOGY

The procedure has been described previously (Budowle and Baechtel 1989). The DNA isolation is described by Baechtel in a separate manuscript in this proceedings (Baechtel 1990) and therefore this protocol will begin with restriction endonuclease digestion.

DNA Restriction

Purified DNA is restriction digested using *Hae*III (Bethesda Research Laboratory, Gaithersburg, MD) according to the manufacturer's specifications. Five units of enzyme are used for each microgram of genomic DNA. After digestion, the DNA is precipitated with ethanol so an appropriate quantity of DNA can be subjected to electrophoresis.

Gel Electrophoresis

Submarine gel electrophoresis is performed in 1.0% agarose gels (Seakem ME) (11 X 14 cm) (6-10 mm thickness) in tris-acetate-EDTA buffer containing ethidium bromide (0.5 mg/L) as described by Maniatis *et al.* (1982). The DNA samples are mixed in a 3:1 or 4:1 ratio with loading buffer (50% glycerol, 0.1% bromophenol blue, 0.1 M EDTA, all in TE buffer) (TE = 10 mM tris and 0.1 mM EDTA, pH 7.5). The samples are applied to wells in the cathodal end of the gel. Electrophoresis is carried out in a BRL H5 submarine tank (Bethesda Research Laboratory, Gaithersburg, MD) at 32 V/cm at ambient temperature for 16 hours or until the dye front is one cm from the anodal end of the gel.

Southern Blotting

The transfer technique is a composite of procedures reported by Southern (1975), Wahl *et al.* (1979) and Reed and Mann (1985). After electrophoresis, the DNA in the gel is denatured by washing in 0.4 M NaOH for 30 minutes. The DNA is transferred in 0.4 M NaOH to Zeta-Probe nylon membrane (Bio-Rad). The blotting apparatus is similar to that described by Gatti *et al.* (1984) (Feinberg and Vogelstein 1983). Blotting time is six hours at ambient temperature. After the DNA is transferred by capillary action, the membrane is washed briefly in a solution containing 2XSSC (20XSSC =175.3 g NaCl and 88.2 g sodium citrate/liter, pH 7.0) and 0.2 M tris, pH 7.5, to remove residual agarose. The membrane(s) is placed between Whatmann 1 MM paper and baked in a vacuum oven at 80° C for 30 minutes.

Probes

The probes used hybridize with VNTRs, each residing at a specific and separate locus. They are YNH24 (D2S44) (Nakamura *et al.* 1987b), CMM101 (D14S13) (Nakamura *et al.* 1988a), 3'HVR (D16S85) (Wainscoat *et al.* 1987), V1 (D17S79) (Balazs *et al.* 1989), pH 3.0 (D4S139) (Milner *et al.* 1989), MS1 (D1S7) (Wong *et al.* 1987), pR365 and L355 (Dennis-Keller *et al.* 1987).

Probe Labeling

Random primer labeling is accomplished according to the manufacturer's instructions contained within the BRL Random Primer DNA Labeling System Kit or according to the method of Feinberg and Vogelstein (Feinberg and Vogelstein 1983, 1984).

Hybridization

Following heat baking, the membrane is placed directly into the hybridization solution, omitting a prehybridization wash. The hybridization solution contains 10% polyethylene glycol (PEG) (Sigma, MW = 8000) (Amasino 1986), and 7% sodium dodecylsulfate (SDS) (Bio-Rad) (Church and Gilbert 1984) in 1.5XSSPE (20XSSPE = 174 g NaCl, 27.6 g $NaH_2PO_4.H_2O$ and 7.4 g EDTA/liter, pH 7.4 (Reed and Mann 1985).

Up to six membranes (11 X 12.5 cm) can be placed in an appropriate-sized plastic sandwich box containing 60 ml of hybridization solution. Once the membrane(s) has been placed in the hybridization solution, labeled probe (5 X 10^5 or 1 X 10^6 dpm/ml of hybridization solution; 2.5 X 10^5 dpm/

ml for pH 3.0) and 1.5 ml of herring sperm DNA (10 mg/ml) (15 mg/60 ml of hybridization solution) are boiled in 0.6 ml of 0.2 M NaOH for five minutes. The denatured probe and herring sperm DNA are added immediately to the hybridization solution. Hybridization is carried out overnight at 65° C with gentle agitation in a New Brunswick Scientific Incubator Series 25 Shaker.

Stringency Washes

After hybridization, the membrane(s) is washed two times in a solution containing 2XSSC and 0.1% SDS for 15 minutes each at ambient temperature. The length of time for the high stringency wash (0.1XSSC and 0.1% SDS at 65° C) varies depending on the particular probe used. For example, probe YNH24 (D2S44) requires a 10 minute high-stringency wash while probe CMM101 (D14S13) requires a 30 minute wash.

Autoradiography

After appropriate stringency washing, the membrane is blotted briefly between Whatmann 1 MM papers. While still moist, the membrane is wrapped in Glad Cling Wrap. The membrane then is sandwiched between two Kodak X-OMAT X-ray films and this complex is flanked with Dupont Cronex Lighting Plus Intensifying Screens. Film exposure is carried out at -80° C. After appropriate exposure time, the film is developed in a Kodak M35A X-OMAT Processor.

DISCUSSION

Although the aspects of this protocol have been addressed previously (Budowle and Baechtel 1989), a few considerations are worth reiterating. These issues are the choice of restriction endonuclease, prehybridization, components of the hybridization solution, and overall sensitivity of detection.

Restriction Endonuclease

Class II restriction endonucleases (REs) cleave DNA at a particular recognition sequence. These enzymes have been used to detect polymorphisms in the DNA sequence. The term RFLP was introduced to describe polymorphisms due to the creation or elimination of a recognition site and more recently due to variation of the length of DNA between RE recognition sites (that is, VNTRs) (Wyman and White, 1980; Nakamura *et al.* 1987a; Wong *et al.* 1987).

There are many REs that can be used for RFLP analysis; however, several factors should be

considered when choosing a RE for forensic applications. They are:

1) The RE should cleave in the flanking sequences of several independent and highly polymorphic VNTR loci.
2) The sizes of generated RFLPs should be amenable to conventional submarine gel electrophoresis.
3) The RE should function with no adverse effects over a wide range of conditions.
4) The RE should be insensitive to mammalian methylation patterns.
5) The RE should be relatively inexpensive.

Several REs, particularly *Hinf*I (Gill *et al.* 1985) and PstI (Balazs *et al.* 1989), have been used for RFLP analysis of forensic biospecimens. Population data was available for some VNTR loci using these REs. However, the limited available population data was not a compelling argument to adopt either *Hinf*I or *Pst*I. It was more important to satisfy the above criteria for forensic analysis. Since the RE - *Hae*III - satisfies these criteria, it is the recommended RE for RFLP analysis of forensic biological materials. *Hae*III recognizes the four base sequence G-G-C-C and forms a blunt end cut between the G and C (Roberts 1987). Since PstI recognizes a six base sequence (C-T-G-C-A-G), *Hae*III generated DNA fragments are smaller than those produced using *Pst*I (Budowle *et al.* 1990). Genetic differences due to the number of repeat sequences of a VNTR locus are more likely to be resolved by electrophoresis when the DNA fragments are small. Additionally, a comparison of *Hae*III, *Hinf*I (recognizes G-A-N-T-C) and AluI (recognizes A-G-C-T) was made regarding fragment size on the D2S44 locus and it was found that *Hae*III yielded the smallest fragments (approximately 1,100 bases less than *Hinf*I and 650 bases less than *Alu*I) (Budowle *et al.* 1990). Thus, superior resolution of the D2S44 variants was achieved.

The laboratories of the RCMP and FBI have found that *Hae*III is compatible with not only D2S44 (Nakamura *et al.* 1987b) but also D1S7 (Wong *et al.* 1987), D4S139 (Milner *et al.* 1989) D14S13 (Nakamura *et al.* 1988a), D16S85 (Wainscoat *et al.* 1989), D17S79 (Balazs *et al.* 1989), D17S74 (Nakamura *et al.* 1988b) and D20S15 (Donnis-Keller *et al.* 1987). These are all highly polymorphic loci and their compatible probes provide a good sensitivity of detection (Budowle and Baechtel 1989). It is anticipated that many other VNTR systems will be found to be useful with *Hae*III with the ancillary benefit of improved reso-

lution of the RFLP patterns (as observed with D2S44).

Methylation patterns can vary among DNA in the various tissues of an individual (Silva and White 1988; Adams and Burdon 1985; Cooper 1983; Ehrlich and Wayne 1981). Thus, REs that are sensitive to methylation of their recognition sites may generate different profiles for different tissues of the same individual. This is of paramount concern for RFLP analysis of evidentiary materials, since often a comparison of DNA profiles derived from different tissues is made. The most common methylation pattern in mammals in the methylation of a cytosine residue followed immediately by a guanine residue (Adams and Burdon 1985; Cooper 1983; Ehrlich and Wayne 1981; Razin and Riggs 1988). A CpG dinucleotide is not in the restriction site for *Hae*III; and if the 3' cytosine in the restriction site is methylated there is no effect on *Hae*III catalytic activity (whereas the activity of *Hinf*I is affected by methylation at the 3' end of its recognition site (Budowle *et al.* 1990; Huang *et al.* 1982; Pech *et al.* 1979; Nelson and McClelland 1989). There have been no observable differences in RFLP patterns generated using *Hae*III from a variety of tissue sources derived from the same individual.

A very important consideration when selecting a RE is that it functions faithfully over a wide range of conditions. Samples encountered at crime scenes are expected to be contaminated. Some contaminants will affect the ability of a RE to digest the DNA in an evidentiary sample such that no digestion or partial digestion might occur. Both Blakesley *et al.* (1977) and Budowle *et al.* (1990) have demonstrated that *Hae*III can properly digest DNA under a variety of ionic strengths, temperature and pH conditions. Thus, *Hae*III should be able to properly digest human genomic DNA that contains cellular and environmental contaminants more readily than less robust REs.

In summary, the selection of *Hae*III for RFLP analysis of evidentiary materials was based on the following factors: *Hae*III is a four base cutter; it provides small resolvable DNA fragments for the D2S44 locus; it is compatible (that is, not cleave the VNTR sequences themselves) with several independent VNTR loci; it is insensitive to the normal methylation patterns that are encountered in mammals; and it is sufficiently resilient to digest potentially contaminated DNA samples. Finally, although prices for REs are subject to change, *Hae*III currently is an inexpensive enzyme.

Prehybridization

After Southern Blotting (Southern 1975) and prior to hybridization of the probe with the immobilized DNA on the membrane, the unoccupied sites on the membrane need to be blocked. This step was necessary to eliminate sites on the membrane where the probe may bind nonspecifically and increase background staining. The time for prehybridization generally ranged from two hours to overnight (Maniatis *et al.* 1982; Wahl *et al.* 1979; Gatti *et al.* 1984). The conditions described in this protocol enable the elimination of the prehybridization step. The membrane can be placed directly in the hybridization solution and labeled probe can be added immediately. Thus, assay time is reduced by one day without loss of sensitivity of detection and without increase in background staining (Budowle and Baechtel 1989). The prehybridization step was eliminated by using 7% SDS in the hybridization solution (Budowle and Baechtel 1989; Church and Gilbert 1984). The high concentration of SDS appears to block instantaneously unoccupied sites on the membrane.

Hybridization Solution

Many components historically used in a hybridization solution appear to be unnecessary. Reagents such as formamide (Kourilsky 1974), Denhardt's solution (Denhardt 1966), milk (or blotto) (Johnson *et al.* 1984) and dextran sulfate (Wahl *et al.* 1979) can be eliminated or replaced with less costly and more effective components. The hybridization solution contains only 7% SDS (Church and Gilbert 1984), 10% polyethylene glycol (PEG) (MW = 8000) (Amasino 1986), and phosphate buffer (Reed and Mann 1985). The advantage of using SDS as a means to eliminate the prehybridization step has been discussed. An additional benefit of a high concentration of SDS in the hybridization solution is an increase in the sensitivity of detection of the system. Whether the effect is due to that suggested for dextran sulfate (that is, probe networking) (Wahl *et al.* 1979), an improvement in hybridization kinetics, or some unexplained cause remains to be determined. The procedure has a sensitivity of detection such that quantities as little as 10 ng of human genomic DNA are sufficient to provide a RFLP pattern. Of course, the use of VNTR probes also contributes to the increased sensitivity of detection for RFLP analysis (Wong *et al.* 1987; Budowle and Baechtel 1989). RFLP profiles can be obtained from samples with as little as 10-50 ng of DNA.

The effect, obtained using dextran sulfate in the hybridization solution (Wahl *et al.* 1979), can be achieved with PEG, but for substantially less cost (Budowle and Baechtel 1989; Amasino 1986). Additionally, PEG can be solubilized more readily than dextran sulfate. Formamide was used in hybridization solutions to reduce the Tm (generally to 42° C for human genomic DNA) for probe:target duplex formation. Temperatures as high as 65° C have a deleterious effect on nitrocellulose membranes; however, nylon membranes, which are used routinely now are more resilient and can withstand temperatures of 65° C for hybridization. Formamide was eliminated as a component of the hybridization solution because it affects hybridization kinetics (Britten *et al.* 1974; Southern 1979) and can reduce the sensitivity of detection of the RFLP system (Budowle and Baechtel 1989). Thus, the efficacy of the system was augmented for less cost. Other reagents which have been used in hybridization solutions, such as Denhardt's and milk, offer no improvement to the efficacy of the analysis (Budowle and Baechtel 1989).

Finally, a simplified hybridization solution with few components has the additional benefit of being much easier to trouble-shoot defective reagents, compared with more complex hybridization solutions.

CONCLUSION

A procedure has been provided that is reliable, simple, reduced in cost, less tedious and efficacious. This method has been found to be applicable to the analysis of evidentiary samples, such as blood, semen, saliva and muscle tissue. It is anticipated that such a technique can be transferred to the forensic laboratory and used on a routine basis.

REFERENCES

Adams, R. L. and Burdon, R. H. (1985). Molecular Biology of DNA Methylation, (Rich, A. ed.), Springer Verlag, NY, pp. 43–333).

Amasino, R. M. (1986). Acceleration of nucleic acid hybridization rate by polyethylene glycol, Anal. Biochem. 152:304–307.

Baechtel, F. S. (1990). The Extraction, Purification and Quantification of DNA, In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U. S. Government Printing Office, Washington, D.C., pp. 25–28.

Balazs, I., Baird, M., Clyne, M. and Meade, E. (1989). Human population genetic studies of

five hypervariable DNA loci, Amer. J. Hum. Genet. 44:182–190.

Blakesley, R. W., Dodgson, J. B., Nes, I. F. and Wells, R. D., (1977). Duplex regions in single-stranded phi X174 DNA are cleaved by a restriction endonuclease from Haemophilus aegyptius, J. Biol. Chem. 252:7300–7306.

Britten, R. J., Graham, D. E. and Neufeld, B. R. (1974). Analysis of repeating DNA sequences by reassociation, Meth. Enzymol. 29:363–418.

Budowle, B. and Baechtel, F. S. (1989). Modifications to improve the effectiveness of restriction fragment length polymorphism typing, Appl. Theor. Electro. (in press).

Budowle, B., Waye, J. S., Shutler, G. G and Baechtel, F. S. (1990). HaeIII - a suitable restriction endonuclease for restriction fragment length polymorphism analysis of biological evidence samples, J. Forens. Sci. 35:(in press).

Church, G. M. and Gilbert, W. (1984). Genomic sequencing, Proc. Natl. Acad. Sci. USA 81:1991–1995.

Cooper, D. N. (1983). Eukaryotic DNA methylation, Human Genetics 64:315–333.

Denhardt, D. T. (1966). A membrane-filter technique for the detection of complementary DNA, Biochem. Biophys. Res. Comm. 23:641–646.

Donnis-Keller, H., Green, P., Helms, C., Cartinhour, S., Weiffenbach, B., et al. (1987). A genetic linkage map of the human genome, Cell 51:319–337.

Ehrlich, M. and Wang, R. Y. (1981). 5- Methylytosine in eukaryotic DNA, Science 212:1350–1357.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity, Anal. Biochem. 132:6–13.

Feinberg, A. P. and Vogelstein, B. (1984). Addendum. Anal. Biochem. 137:266–267.

Gatti, R. A., Concannon, P. and Salser, W. (1984). Multiple use of Southern Blots, BioTechniques 2:148–155.

Gill, P. Jeffreys, A. J. and Werrett, D. J. (1985). Forensic Application of DNA fingerprints, Nature 318:577–579.

Gillespie, D. and Spiegelman, S. (1985). A quantitative assay for DNA-RNA hybrids with DNA immobilized on a membrane, J. Mol. Biol. 12:829–842.

Huang, L. H., Farnet, C. M., Ehrlich, K. C. and Ehrlich, M. (1982). Digestion of highly modified bacteriophage DNA by restriction endonucleases, Nuc. Acids Res. 10:1579–1591.

Johnson, D. A., Gautsch, J. W., Sportsman, J. R. and Elder, J. H., (1984). Improved technique utilizing nonfat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose, Gene Anal. Techn. 1:3–8.

Kourilsky, P. Mercereau, O. and Tremblay, G. (1974). Hybridization on filters with complete DNA in the liquid phase in a standard and micro assay, Biochimie 56:1215–1222.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). Molecular Cloning. A laboratory manual. Cold Spring Harbor Laboratory Publication, NY, pp. 438–454.

Milner, E. C., Latshaw, C. L., Willems van Dijk, Charmely, P., Concannon, P. and Schroeder, H. W. (1989). Isolation and mapping of a polymorphic DNA sequence (pH30) from human chromosome 4 (D4S139), Nuc. Acids Res. 17:4002.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987a). Variable number of tandem repeat markers for human gene mapping, Science 235:1616–1622.

Nakamura, Y., Gillilan, S., O'Connell, P., Leppert, M., Lathrop, G. M., Lalouel, J. M. and White, R. (1987b). Isolation and mapping of a polymorphic DNA sequence pYNH24 on chromosome 2(D2S44), Nuc. Acids Res. 15:10073.

Nakamura, Y., Culver, M., Gill, J., O'Connell, P., Leppert, M., Lathrop, G., Lalouel, J. M. and White, R., (1988a). Isolation and mapping of a polymorphic DNA sequence pMLJ14 on chromosome 14 (D14S13), Nuc. Acids Res. 16:381.

Nakamura, Y., Martin, C., Myers, R., Ballard, L., Leppert, M., O'Connel, P., Lathrop, G., Lalouel, J. and White, R. (1988b). Isolation and mapping of a polymorphic DNA sequence (pCMM86) on chromosome 17q (D17S74), Nuc. Acids Res. 16:5223.

Nelson, M. and McClelland, M. (1989). The effect of site-specific methylation on DNA modification methyltransferases and restriction endonucleases, Nuc. Acids Res. 17:r389–r415.

Odelberg, S. J., Plaetke, R., Eldridge, J. R., Ballard, P., O'Connel, P., Nakamura, Y., Leppert, M., Lalouel, J. M. and White, R. (1989). Characterization of eight VNTR loci by agarose gel electrophoresis, Genomics 5:915–924.

Pech, M., Streeck, R. E. and Zachau, H. G. (1979). Patchwork structure of a bovine satellite DNA, Cell 18:883–893.

Razin, A. and Riggs, A. D. (1980). DNA methylation and gene function, Science 210:604–610.

Reed, K. C. and Mann, D. A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes, Nuc. Acids Res. 13:7207–7221.

Roberts, R. J. (1987). Restriction enzymes and their isoschizomers, Nuc. Acids. Res. 15:(Suppl.) r189–r217.

Silva, A. J. and White, R. (1988). Inheritance of allelic blueprints for methylation patterns, Cell 54:145–152.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis, J. Mol. Biol. 98:503–517.

Southern, E. M. (1979). Gel electrophoresis of restriction fragments, Meth. Enzymol. 68:152–176.

Wainscoat, J. S., Pilkington, S., Peto, T. E. A., Bell, J. I. and Higgs, D. R. (1987). Allele-specific DNA identity patterns, Human Genetics 75:384–387.

Wahl, G. M., Stern, M. and Stark, G. R. (1979). Efficient transfer of large DNA fragments form agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate, Proc. Natl. Acad. Sci. USA 76:3683–3687.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. J. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA, Ann. Hum. Genet. 51:264–288.

Wyman, A. R. and White, R. (1980). A highly polymorphic locus in human DNA, Proc. Natl. Acad. Sci. USA 77:6754–6758.

STRATEGIES FOR LABELING AND DETECTION OF NUCLEIC ACID PROBES

Terry A. Landers

Life Technologies, Inc.
Gaithersburg, Maryland

The development of molecular probes has created powerful tools with applications in many areas of the biological sciences. These probes are molecules which are capable of finding and reporting on the presence of specific targets that are other molecules and macromolecular structures. Targets may be in solution, in a gel, or fixed to a filter or a microscope slide. Probes are able to report their presence because they carry some distinguishable feature, a label, that allows the user to determine where and/or how much target was found by the probe. The development of probes has been made possible by the development of techniques for labeling macromolecules without significantly impairing their natural function. Antibodies, with radioisotopes or enzymes attached, are the most widely used molecular probes, capable of reporting qualitative and quantitative information on a spectrum of targets ranging from small hormone and drug metabolites to complex viral and cellular antigens. Compared to antibody probes, nucleic acid probes have a narrow range of target types, only other nucleic acids, but within that range they can be designed to recognize an almost infinite number of different nucleic acid sequences.

The most fundamental requirement of a nucleic acid probe is that it contain the complement of the target sequence that is to be detected. Recognition depends on the ability of the probe to form a double helix with the target through Watson-Crick base pairing. Molecules containing the necessary sequences can be derived from natural sources by recombinant DNA cloning techniques, by purification from enriched tissue sources, or by chemical synthesis of short nucleic acid molecules (oligonucleotides) of known sequence. A corollary is that a probe must not contain significant amounts of sequences, such as cloning vector, that might detect other unintended targets.

To be useful as a nucleic acid probe, a nucleic acid molecule must be modified in some way to be detectable in the presence of a large excess of non-target nucleic acid. An average size human gene may constitute less than one part per million of cellular DNA and many of the targets of forensic interest are even smaller. To compound the detection problem, the total amount of DNA recovered from a sample is often on the order of 0.1 to 10 μg, the amount of DNA contained in a few thousand cells. Thus, the mass of actual target may be no more than a fraction of a picogram (10^{-12} g) to which a similar quantity of the probe will hybridize. At present, the most reliable means for detecting the smallest amounts of target require the use of radioisotopes, most commonly P^{32}. Non-radioactive detection methods have also been developed and are commonly used for detection of more abundant targets. Some chemiluminescent methods are now reaching the level of sensitivity of P^{32}, as will be discussed later.

METHODS OF LABELING NUCLEIC ACIDS

Labels can be introduced into a nucleic acid molecule by enzymatic, non-enzymatic or a combination of enzymatic and non-enzymatic methods. The enzymatic methods are generally the most widely used because they can generate either DNA or RNA probes of a wide range of sizes, the reactions are easily controlled, the products are well defined, and the process is relatively inexpensive. Some labels are attached to the sugar, others to the nitrogenous base of nucleotides. The term labeled base is often used as a synonym for labeled nucleotide without meaning that the label is necessarily in the base portion of a nucleotide. Base and nucleotide are often used interchangeably.

Enzymatic Labeling Methods

Nick translation: Nick translation (Rigby *et al.* 1977; Figure 1), a widely-used technique, is based on the properties of DNA polymerase I (Pol I) from *E. coli.* This DNA polymerase requires a template, a single-stranded region of DNA for which a complementary strand will be synthesized, and a primer, a nucleic acid hybridized to the template, with a free 3'-OH group, to which the newly synthesized complimentary DNA strand will be attached. The new DNA is synthesized from deoxyribonucleoside triphosphates (dNTP's). DNA synthesis proceeds by attaching the deoxyribonucleoside monophosphate portion of a dNTP to the 3'-end of the primer and releasing pyrophosphate. The choice of base attached is dictated by the op-

posing base in the template and the base-pairing rules. The new strand grows in a 5′ to 3′ direction.

In addition to its DNA polymerase activity, Pol I also has two exonuclease activities, one capable of removing nucleotides from the 3′ end, the other removing them from the 5′ end of double-stranded DNA. The 3′-exonuclease serves *in vivo* as a proof reading activity, removing incorrectly incorporated nucleotides from newly synthesized DNA. The 5′-exonuclease functions *in vivo* to remove damaged bases during one type of repair

process. It is this repair capability that is used for nick translation.

To perform a nick translation labeling reaction, the double-stranded DNA to be labeled is mixed with a small amount of pancreatic DNAse I, Pol I and a mixture of dNTP's, at least one of which is labeled. Pancreatic DNAse I introduces nicks into the probe DNA by cleaving the phosphate-sugar backbone of either strand of the molecule at random. These nicks are recognized by Pol I as sites for repair. Pol I binds to a nick and its 5′-exonuclease activity removes the nucleotide on the
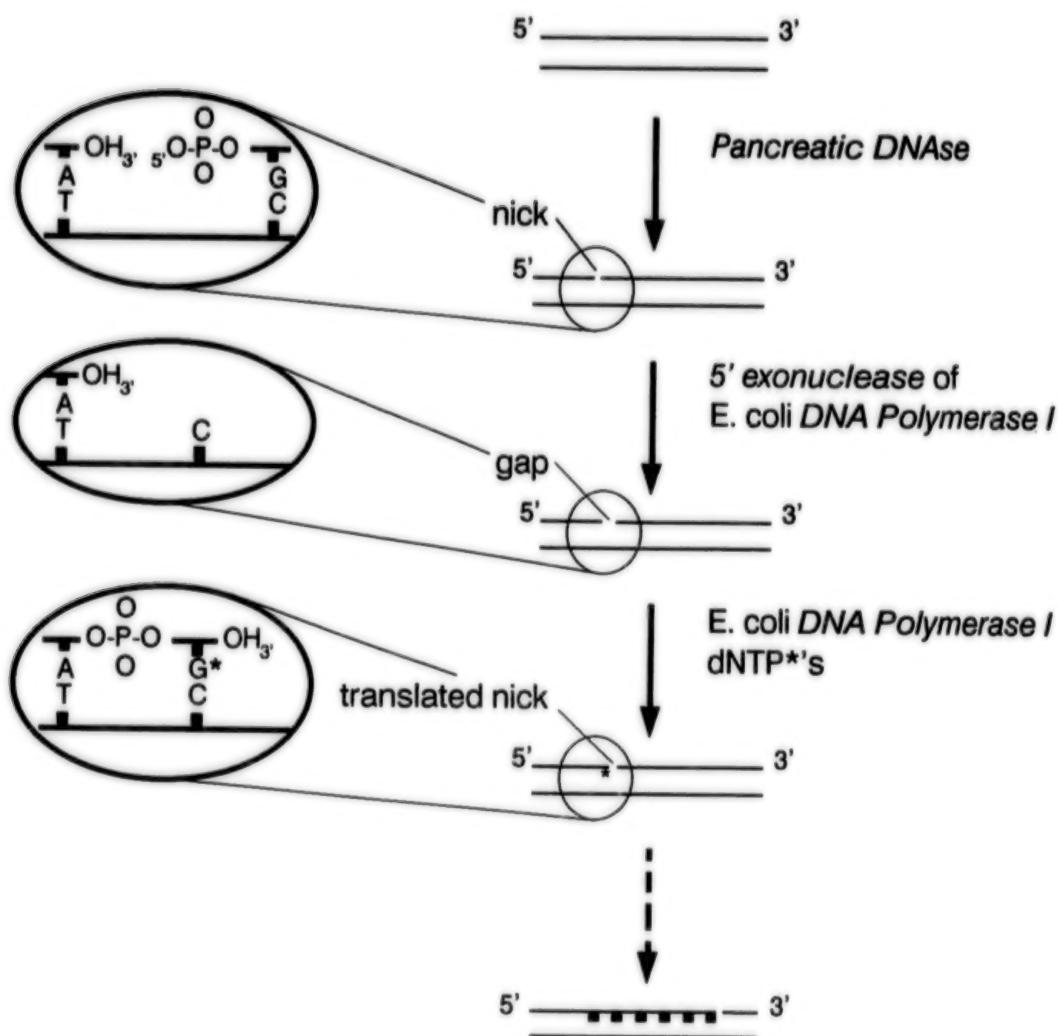

Figure 1. Nick translation. Labeled nucleotides are incorporated into double stranded DNA by the combined action of pancreatic DNase I and the 5′-exonuclease and polymerase activities of *E. coli* DNA polymerase I.

5'-side, leaving a one-nucleotide gap. The polymerase activity then repairs the gap, incorporating a new nucleotide from the pool of labeled dNTP's by attaching it to the free 3'-OH side of the gap. This process still leaves a nick, but it has now been translated one nucleotide in the 3' direction from its original location. This process is repeated until the nick is translated to the end of the template, meets another nick on the opposite strand, or the Pol I spontaneously releases itself from the nick.

Since DNAse I nicks both strands of DNA at random, labeled nucleotides are evenly distributed over both strands of all DNA in the reaction mixture. The final labeled DNA product is no longer continuous, full length double strands due to the introduction of the nicks. When the starting DNA is relatively large, the introduction of nicks does not affect the utility of the probe, but this method is not useful for labeling DNAs less than ~500 bases long as it leaves many fragments too small to hybridize efficiently.

Random oligonucleotide primers: Another method for labeling probes uses a mixture of short oligonucleotides (average length 6 bases) of random sequence to prime DNA synthesis on single-stranded DNA molecules (Feinberg and Vogelstein 1983, 1984; Figure 2). DNA synthesis is carried out by a modified form of *E. coli* DNA polymerase I. Pol I is modified by enzymatic or genetic methods to remove the 5'-exonuclease activity which would otherwise degrade the oligonucleotide primers before they could be used. This modified form is referred to variously as Klenow, Klenow fragment, DNA Polymerase I Large Fragment or Pol IK.
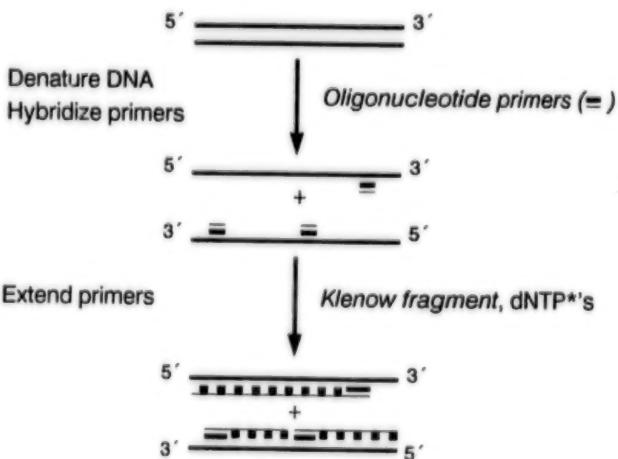
Figure 2. Random oligonucleotide primers. Oligonucleotides hybridize to single-stranded templates to provide primers for synthesis of labeled DNA by the Klenow fragment of *E. coli* DNA polymerase I.

For oligonucleotide priming, the template DNA must be single stranded. Double-stranded DNA's can be denatured into the two complementary single strands either by heating to >95° C then cooling quickly or by raising the pH to >11.5 then neutralizing quickly. Once the template is single stranded, the oligonucleotide primers, Klenow fragment and labeled dNTP's are added. The short, random-sequence primers have a high probability of hybridizing to many places on the template DNA. When a Klenow fragment finds a hybridized primer, the polymerase activity then begins to synthesize a complementary strand from the position of the primer. The newly synthesized DNA is attached to the primer and includes label from the pool of labeled dNTP's. If the Klenow fragment encounters another primer while elongating, the second primer is displaced so elongation can continue. Extension stops when the enzyme encounters the end of the template or another extended double-stranded stretch, or the enzyme spontaneously terminates. This technique works well for DNA molecules of almost any size, including those too small to label by nick translation. Since this procedure results in a net synthesis of new DNA (potentially one new strand for each template strand), rather than replacement of DNA as with nick translation, incorporation of label is usually more efficient. This procedure also benefits from not being dependent on the sometimes temperamental pancreatic DNase.

A variant of this method uses a specific, rather than random, primer to direct synthesis from only a specified region of one strand of a DNA template. This specific primer reaction is basically the same as that used for DNA sequencing by the dideoxy method of Sanger *et al.* (1977). The specific primers are generally synthetic oligonucleotides 12 to 20 bases long. With a carefully chosen primer, little or no probe complementary to any contaminating DNA, such as residual cloning vector, will be synthesized. This can be a problem with probes labeled by nick translation or with random primers because these methods label all DNA's in a solution, including residual cloning vector fragments. Labeled vector DNA may detect related sequences in samples containing contaminating bacterial DNA, giving false bands or high backgrounds.

T4 DNA polymerase replacement synthesis: T4 DNA polymerase is produced by bacteriophage T4 after infection of *E. coli.* Enzymatically it is similar to the Klenow fragment in that it has both DNA polymerase and 3'-exonuclease activities but lacks a 5'-exonuclease activity. The principal difference be-

tween the two is the rate at which the 3' exonuclease acts. In the absence of dNTP's to polymerize, T4 DNA polymerase has a much more active 3' exonuclease activity. It will rapidly remove bases from the 3' ends of all DNA molecules in a solution, shortening all molecules synchronously. It is this capability that led to its use in replacement synthesis labeling procedures (O'Farrell 1981; Deen *et al.* 1983).

In practice (Figure 3), T4 DNA polymerase is added to the double-stranded probe DNA in the absence of dNTP's for a length of time adequate to allow the 3' exonuclease to resect the molecules to just over half their length. There section process exposes stretches of single-stranded DNA, which will later serve as templates, and shortened strands with free 3' ends, which will later serve as primers. Since resection proceeds from both ends of a molecule, care must be taken that it does not reach the middle of the molecule or the primer strands will be too short to overlap.
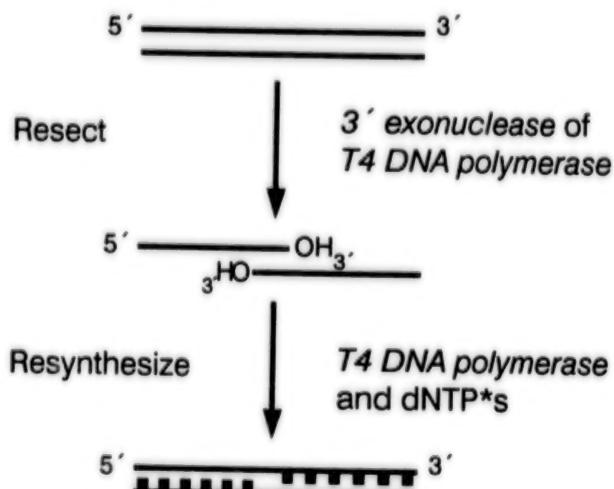
Figure 3. T4 DNA polymerase replacement synthesis. The 3'-exonuclease activity of bacteriophage T4 DNA polymerase removes nucleotides from the 3' end of DNA molecules in the absence of dNTP's. Upon addition of labeled deoxynucleoside triphosphates, the polymerase synthesizes a labeled replacement of the DNA previously removed.

After resection, labeled dNTP's are added to the reaction and synthesis of new, labeled DNA begins on the resected molecules. Although the 3' exonuclease is still active, the polymerase is capable of adding nucleotides faster than they are removed by the exonuclease. Given sufficient dNTP's, the product is mostly full length, double-stranded DNA, although each individual strand is labelled only in its 3' half, and a region in the center of the molecule is not labeled. This asymmetry of labeling does not affect the utility of such probes for most uses. The efficiency of labeling is

similar to that obtained by nick translation. With careful control of resection times, this method can be used on double-stranded DNAs ranging in length from <100 base pairs (bp) to >20,000 bp.
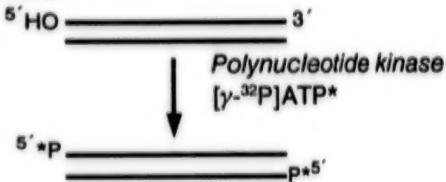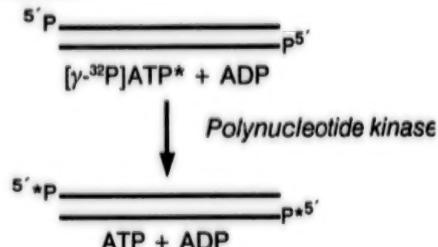
Figure 4. Polynucleotide kinase. a. The forward reaction catalyzed by bacteriophage T4 polynucleotide kinase transfers the radioactive terminal phosphate residue from [γ-P^{32}]ATP to the 5'-hydroxyl group of a polynucleotide. b. Under different reaction conditions, polynucleotide kinase catalyzes an exchange of phosphate groups between [γ-P^{32}]ATP and 5'-phosphate groups on polynucleotides. This reaction also requires ADP.

Polynucleotide kinase: Both single- and double-stranded polynucleotides can be labeled specifically at their 5' ends with P^{32} by T4 polynucleotide kinase. This enzyme, isolated from bacteriophage T4-infected *E. coli,* is capable of either adding or exchanging a monophosphate group at the 5' end of oligo- and polynucleotides with the -phosphate group from ATP (Maniatis *et al.* 1982; Figure 4).

For labeling by polynucleotide kinase, the template must have either a hydroxyl group or a monophosphate at the 5' end. Labeling is most efficient for hydroxyl-terminated molecules such as are commonly found on synthetic oligonucleotides. The 5'-monophosphate ends normally are generated by cleavage of DNA's with restriction endonucleases and other exo- and endonucleases such as pancreatic DNAse I. Monophosphates may be removed from 5'-termini of any DNA or RNA molecule by use of a phosphatase such as *E. coli* alkaline phosphatase or calf intestine alkaline phosphatase.

The only type of label used with polynucleotide kinase is (γP^{32})ATP [the phosphorous in the γ (outermost) phosphate of the triphosphate is radioactive]. In the forward reaction (Figure 4a), polynucleotide kinase transfers the γ phosphate from

ATP to the 5'-hydroxyl group of the polynucleotide, producing a phosphorylated polynucleotide and ADP. Under appropriate conditions, the enzyme is also capable of catalyzing the reverse reaction, transferring a 5'-monophosphate group from a polynucleotide to ADP to produce ATP and a 5'-OH polynucleotide. By selecting conditions where both the forward and reverse reactions can proceed and providing both ADP and (γP^{32})ATP, an exchange reaction occurs with phosphorylated polynucleotides, producing labeled polynucleotides and unlabeled ATP (Figure 4b). The forward reaction is used frequently with high specific activity (γP^{32})ATP to produce highly labeled oligonucleotide probes but it is not an effective way to label large polynucleotides to high specific activity. The exchange reaction is used commonly to label electrophoresis standards where high specific activity is not necessary.

A characteristic of polynucleotides labeled by kinase is that when a mixture of molecules is labeled, all molecules receive approximately the same amount of label, regardless of size, since only 5' ends are labeled. This is in contrast to the previously described techniques in which molecules carry an amount of label proportional to their length. This feature of kinase labeling can be used to advantage when the molar quantities, rather than mass, of molecules of different sizes is important.

Terminal transferase: Calf thymus terminal deoxyribonucleotidyl transferase (TdT) is a primer-dependent, template-independent polymerase. Like all known nucleic acid polymerases it elongates by adding nucleotides to the 3' end of a primer, but in this case, without the direction of a template. The nucleotides must be supplied as dNTP's (Roychoudhury and Wu 1980; Kumar *et al.* 1988; Figure 5). If provided with a labeled mixture of dNTP's, TdT will synthesize a tail of random sequence and random length on accessible 3' termini. If only a single type of dNTP is provided, the tails will be a homopolymer.

5' ═══════════ 3'
    ───────────

    ↓ *Terminal deoxynucleotidyl transferase dNTP**

5' ═══════════ 3'
■■■■───────────▪▪▪▪
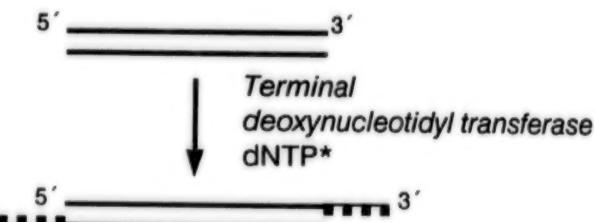
Figure 5. Terminal transferase. Labeled nucleotides are attached to the 3' ends of DNA molecules by terminal deoxynucleotidyl transferase from calf thymus.

The primer for TdT may be either single- or double-stranded DNA. The efficiency of a primer depends on the structure of the 3' end of the molecule. If the 3' end is single-stranded or blunt (base paired exactly), as when cleaved with restriction endonucleases *PstI* or *Hae*III, respectively, nucleotides are added readily. Different buffer conditions are required if the 3' end is recessed (there are unpaired bases at the 5' end of the opposite strand) such as is produced by *Eco*RI or *Hind*III.

The length of the tail can be controlled to some extent by choice of time and reaction conditions. Tail lengths can be restricted to exactly one nucleotide by providing a deoxyribonucleotide analog lacking a 3'-hydroxyl group. Such analogs include the 2',3'-dideoxynucleoside triphosphates used for DNA sequencing and 3'-dATP (cordycepin). Molecules with labeled, single-nucleotide 3' tails are similar to those labeled using the polynucleotide kinase procedure in that only one label is attached per molecule. In addition, TdT the possibility of attaching a single residue of a variety of types of modified nucleotides.

Bacteriophage RNA polymerases: The bacteriophages SP6, T7 and T3 each produces a phage-specific RNA polymerase upon infection of its host. Each phage RNA polymerase recognizes highly specific promoter sequences as the points at which to begin transcription. Normally found only in phage DNA at the beginning of genes, these promoters indicate the strand of DNA and the point at which synthesis of an RNA copy of that strand of the DNA is to begin.

Phage RNA polymerases can be used *in vitro* to produce RNA copies of selected regions of DNA molecules by constructing appropriate promoter-containing recombinant DNA molecules (Green *et al.* 1983; McCracken 1985). The probe sequence to be transcribed is inserted at a site downstream from a phage promoter in a specially designed cloning vector. The cloned DNA is isolated from bacteria as a circular molecule. For *in vitro* work, the template is usually linearized by cleaving the probe DNA at the promoter-distal end of the probe segment with a restriction endonuclease. On properly prepared linear molecules, the RNA polymerase runs off the end after transcribing the probe region and does not transcribe vector sequences to any significant degree. With an adequate supply of NTP's, the phage RNA polymerases will transcribe a single template molecule many times, yielding a 5- to 20-fold net synthesis of probe transcript over input template (Figure 6).
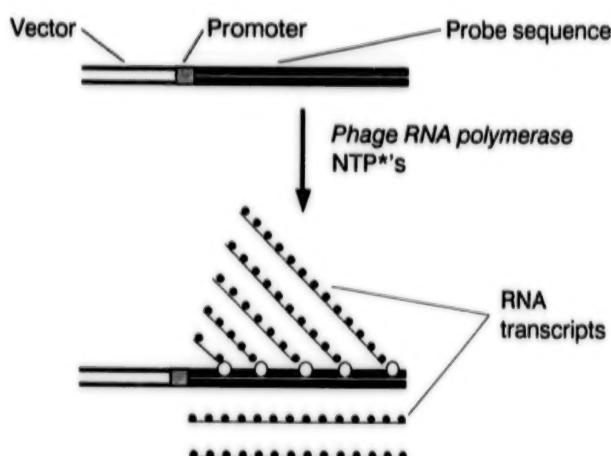
Figure 6. RNA polymerases. Probe sequences cloned into special vectors can be transcribed into labeled RNA by bacteriophage RNA polymerases which initiate RNA synthesis at phage-specific promoter sequences.

The RNA probes can replace DNA probes in most procedures with only minor changes. The RNA probes do require more care in handling at certain stages because they are so sensitive to ribonucleases which are frequent reagent contaminants. This sensitivity can be used to advantage after hybridization to a blot because RNA in a DNA-RNA hybrid is protected from common ribonucleases. Thus, treating an RNA-probed blot with ribonuclease reduces backgrounds due to unhybridized probes. Further, RNA probes have an additional advantage over most DNA probes in that they are naturally single stranded. The absence of a complementary strand eliminates the self-hybridization that competes with target hybridization when double-stranded probes are used.

Non-enzymatic Labeling Methods

A variety of non-enzymatic methods for labeling DNA and RNA have been developed (Table 1). All involve attaching a recognizable substituent to an existing unlabeled nucleic acid, most frequently through covalent bonds to the nitrogenous bases of the nucleotides.

Table 1. RADIOISOTOPE LABEL CHARACTERISTICS

Isotope	Relative Energy	Halflife	Method of Incorporation
P^{32}	high	14 d	enzymatic
S^{35}	moderate	89 d	enzymatic
I^{125}	moderate	60 d	enzymatic or chemical
H^{3}	low	12 y	enzymatic

Photoactivated compounds: Light-generated free radicals are capable of rapidly forming new covalent bonds by insertion into many types of existing covalent bonds. This technique can be used for labeling nucleic acids (and proteins) with the vitamin/cofactor biotin by irradiating a mixture of the nucleic acid to be labeled and biotin linked to an arylazide (Photobiotin™; Forster *et al.* 1985; Figure 7). The modified benzene readily generates anitrene radical when exposed to a common sunlamp. The photoactivated compound then inserts itself into other molecules it encounters in solution. Detection of photobiotinylated probes is generally slightly less sensitive than for enzymatically biotinylated probes.

Figure 7. Photobiotin. The photosensitive arylazide group of photobiotin is activated by exposure to light. The activated molecule then forms covalent bonds with nucleic acids in the solution to produce biotin-labeled probes.

Alkylating compounds: A wide variety of compounds are capable of alkylating the bases of DNA and RNA. Many of these chemicals are also carcinogens. N-acetoxy-N-acetyl-2-aminofluorene (AAF), which alkylates predominantly G residues and was originally studied as a carcinogen, has been used as a label for DNA detection (Tchen *et al.* 1984).

Iodination: Direct oxidative iodination of nucleic acids to generate iodo-C (and iodo-U in RNA) residues has been used for a number of years as a way of introducing I^{125} or I^{131} into unlabeled molecules (Piatyszek *et al.* 1988). Iodo-C and iodo-U are also precursors for further chemical modification. The iodine linkage is not as stable as other modifications and exhibits a slow loss of label under some conditions.

Amino substitution: A two-step procedure for introducing amino-containing substituents to nucleic acids has been reported recently (Keller *et al.* 1988). Bromine is introduced into T, G, or C residues under mild oxidative conditions. The bromine is then displaced by an amino group linked to a reporter group such as dinitrophenol. In another reaction, bisulfite catalyzes transamination between the amino group of cytidine and substituted amines.

This method has been used to attach amine-containing linker arms to CTP and dCTP for further modification (Gillam and Tener 1986).

Glutaraldehyde linking of substituents: Glutaraldehyde can be used to crosslink proteins, enzymes and polyamines to nucleic acids in solution in a manner similar to that used for tissue fixation for microscopy (Renz and Kurz 1984; Al-Hakim and Hall 1986). The degree of crosslinking must be carefully controlled to prevent inter-strand crosslinks and inhibition of hybridization efficiency.

Periodate oxidation: The 2',3'-cis hydroxyls at the 3' end of RNA molecules can be oxidized to a pair of aldehyde groups by periodate oxidation under mild conditions. The aldehydes are then targets for addition of amino-containing compounds through a Schiff-base mechanism.

Chemical synthesis: Chemical synthesis of oligonucleotides permits the introduction of modified bases or their precursors at specified positions in the sequence. The modified oligonucleotide may be used directly as a probe (Jablonski *et al.* 1986) or as a primer (Smith *et al.* 1985).

Enzymatic/non-enzymatic Labeling Methods

Some useful labeled nucleotide analogs are poorly or not significantly incorporated into nucleic acids by the appropriate DNA or RNA polymerases, probably due to the bulky or chemical characteristics of the label. Such labels can be incorporated by a two step process if a chemical precursor of the labeled nucleotide is acceptable to an RNA or DNA polymerase. For example, phage RNA polymerases incorporate biotinylated nucleotides more slowly than normal nucleotides, which reduces the size and yield of biotinylated RNAs. This problem can be circumvented by synthesizing the RNA with a precursor, such as allylamine-UTP, which is incorporated as efficiently as normal UTP (McCracken 1985; Anonymous 1987). The allylamine group provides a reactive primary amino function to which biotin can be attached chemically using (ϵ-caproylamidobiotin N-hydroxysuccinimide ester (CAB-NHS ester, Figure 8b). Similarly, fluorescent DNA can be prepared by incorporation of N^6-(6-aminohexyl)dATP (Figure 8a) and subsequent reaction with fluorescein isothiocyanate or fluorescein N-hydroxysuccinimide ester. Proteins have also been linked to amino-substituted nucleic acids using bifunctional crosslinking reagents (Jablonski *et al.* 1986).

a. N^6-aminohexyl dATP
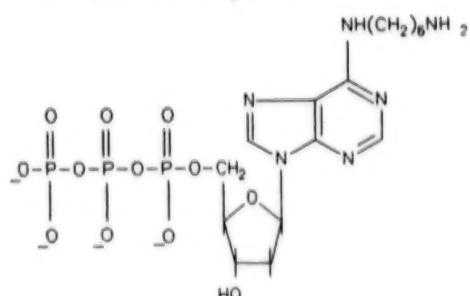
b. ϵ-caproylamidobiotin-N-hydroxysuccinimide ester
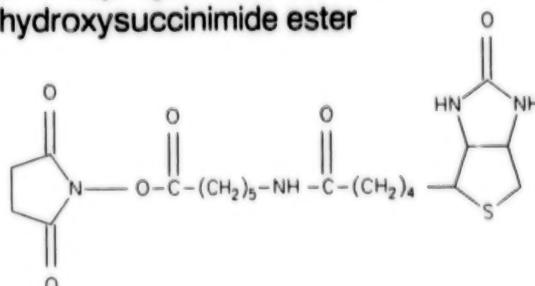
Figure 8. Amine-containing nucleotides. a. N^6-(6-aminohexyl)dATP is a DNA precursor which can be modified by attachment of fluorescein or biotin. c. ϵ-caproylamidobiotin N-hydroxysuccinimide ester (CAB-NHS ester) reacts with primary amines to attach biotin labels to nucleic acids.

LABELS AND DETECTION METHODS

Radioisotope Labels

Selection of isotopes: A number of radioisotopes can be introduced as labels into nucleic acids (Table 2) by either enzymatic or chemical means. The choice of isotope is determined by the source and type of nucleic acid to be labeled, labeling procedures available, desired method of detection and the level of sensitivity needed.

Table 2. NON-ENZYMATIC METHODS FOR LABELING NUCLEIC ACIDS

Photo-activated compounds
Alkylating reagents
Sulfonation
Iodination
Bromination with amino substitution
Glutaraldehyde crosslinking
RNA-periodate oxidation and aldehyde coupling

The most frequently used isotope is P^{32} because of its relatively high energy particle (electron) emission, its high specific activity [disintegrations per minute (dpm) per mole or per microgram] and the fact that phosphorus is a normal constituent of all nucleic acids.

Sulfur (S³⁵) is also a high specific-activity isotope but has a less energetic emission, making common methods of detection less sensitive. Further, S³⁵ is usually incorporated into nucleotides as an oxygen analog which is accepted by most (but not all) enzymes.

Iodine (I¹²⁵) has a specific activity between that of S³⁵ and P³². It emits radiation (photons) of moderate energy. Iodine can be introduced directly into nucleic acids by chemical means (see above) or by coupling to a nucleoside triphosphate followed by enzymatic incorporation of the modified nucleotide.

Tritium (H³) has a relatively low specific activity and a very low energy β emission. Due to the abundance of hydrogen in nucleic acids, many tritium atoms can, in theory, be introduced to give a high specific activity per mole of nucleotide, but the difficulty of detecting the low energy emissions efficiently and conveniently has limited the use of H³ in molecular biology applications.

Detection of Radioisotopes

Autoradiography: The detection of radioactive compounds by their ability to expose photographic film has a long history. The development of specialized films and improved exposure techniques has made autoradiography a very sensitive method. It is currently the most sensitive method available for detection of moderate or strong β emitters on solid substrates such as filters and gels when spatial relationships must be determined.

For direct autoradiography, a dry filter or gel is simply pressed against a sheet of film for a period of time, after which the film is developed to reveal areas exposed by the radioactive emissions. The sensitivity of detection depends on the type of isotope, the type of film and the length of the exposure. When detecting P³² (Figure 9a, right), sensitivity can be enhanced by using a fast film with photographic emulsion on both surfaces. The particle emitted by P³² is energetic enough to travel completely through a sheet of x-ray film and expose silver grains on both surfaces. Additional sensitivity is obtained by the use of an x-ray intensifying screen in a sandwich format. These screens contain compounds which fluoresce or scintillate when struck by x-rays, γ-rays or high energy β particles. The light emitted further exposes the autoradiographic film around the point at which the decay event occurred. The use of double-emulsion films and intensifying screens increases sensitivity by two to five fold at the expense of resolution. Bands on films will be substantially broader when enhancing techniques are used, but for most applications the loss of resolution is not a serious problem. Efficient autoradiographic detection of I¹²⁵ requires the use of intensifying screens because the γ-rays frequently pass through the film without exposing a silver grain.

Direct autoradiography can detect S³⁵ but with significantly less sensitivity than P³². The particle emitted by S³⁵ is not energetic enough to pen-

a. Autoradiography

intensifier screen
film
filter or gel

³⁵S   ³²P

b. Fluorography

film
filter or gel
with scintillant
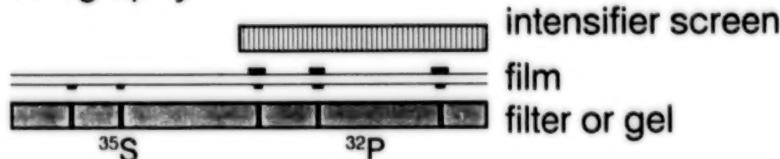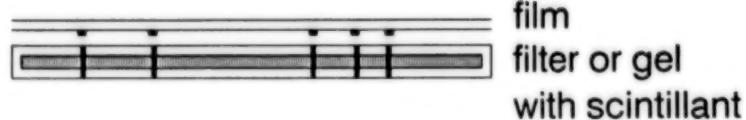
Figure 9. Detection of radioisotopes. a. Autoradiography. Right: detection of P³² is enhanced by use of double emulsion x-ray films and an intensifying screen that emits light when hit by a radioactive decay particle. Left: S³⁵ emissions are not energetic enough to penetrate the plastic base of x-ray film so it only exposes the closest layer of photographic emulsion. b. Fluorography. Impregnating a filter or gel with a radio-sensitive scintillant allows low energy decay events from S³⁵ or H³ to cause a flash of light which exposes the x-ray film.

etrate the plastic base of the film and so can expose only the layer of photographic emulsion facing the filter (Figure 9a, left). Thus, for S^{35}, double-sided films and intensifying screens are of little value. Both P^{32} and S^{35} detection are enhanced by exposure of films at very low ($-70°$ C) temperatures. The low temperature stabilizes partially exposed silver grains. At maximal sensitivity, bands containing less than 0.1 dpm of P^{32} in a band or spot of a few mm² can be detected with an exposure of a few days.

Fluorography: Sensitivity of detection of weaker emitters such as S^{35} and H^3 can be increased by the use of fluorographic techniques (Bonner and Laskey 1974; Roberts 1985). If a gel or filter is soaked in a radio-sensitive scintillant compound, radioactive emissions generate a flash of light that can be detected on x-ray or photographic film (Figure 9b). Fluorography improves film detection of S^{35} approximately 10-fold and H^3 10- to 50-fold for filters. Like intensifying screens, fluorographic detection reduces resolution of closely spaced bands.

Scintillation counting: Reliable detection of low levels of radioisotopes by scintillation counting (liquid scintillation spectrometry) requires accumulation of a statistically significant number of detected emissions above a well-determined background. These conditions can generally be met with liquid-phase and dot blot-type nucleic acid hybridization assays but are difficult to achieve when analyzing Southern blot filters. Since filters must be cut into pieces for such counting, resolution of bands is substantially reduced unless pieces are very small.

Geiger-Mueller counting: The detection of ionizing radiation such as β particles by the Geiger-Mueller (GM) method is not as sensitive as autoradiography or liquid scintillation spectrometry, but with recent improvements may become a valuable technique for some applications. Commercial instruments are capable of scanning filters and accumulating x,y plots of radioactivity with data being stored and analyzed by an associated computer. With appropriate data collection times and analysis, sensitivity (but not resolution) is beginning to approach that of autoradiography.

Non-radioactive Labels

Some non-radioactive labels that are covalently attached to a nucleic acid are able to report their presence directly, analogous to radioisotopes, but many require additional steps or intermediates for detection (Table 3).

Table 3. NON-RADIOACTIVE LABELS USED FOR NUCLEIC ACID DETECTION

Haptens and cofactors
  Biotin
  N-acetoxy-N-acetyl-2-aminofluorene
  Dinitrophenol
  Fluorescein
  Digioxigenin
Enzymes
  Horseradish peroxidase
  Calf intestine alkaline phosphatase

Fluorescent labels: Nucleic acids can be modified to fluoresce either by incorporation of fluorescent nucleotides during enzymatic or chemical synthesis, or by post-synthesis modification such as described above. Addition of large, usually hydrophobic fluorescent groups may have a significant effect on hybridization behavior of the probe. Detection of such probes on filters is often difficult since several of the common filter materials are fluorescent themselves. In addition, colored impurities in some samples may cause variable degrees of fluorescence quenching, particularly in dot blot assays. Fluorescent polynucleotides or conjugates can be detected efficiently when the nucleic acid is in solution, such as during electrophoresis in automated DNA sequencing equipment (Connell *et al.* 1987; Prober *et al.* 1987; Smith *et al.* 1985).

Enzyme-linked probes: Enzymes such as alkaline phosphatase and horseradish peroxidase can be linked to nucleic acids as reporters as is done for antibody-based ELISA's. In a two-step method, a strongly basic protein or polyamine is attached to DNA with glutaraldehyde, then the reporter enzyme is conjugated to the protein-DNA complex (Renz and Kurz 1984). Synthetic oligonucleotides which incorporate a primary amine have been crosslinked directly to alkaline phosphatase with bifunctional crosslinking reagents (Jablonski *et al.* 1986). The reporter enzymes used for this technique must be stable to treatment encountered in hybridization reactions, which may include use of detergents such as sodium dodecylsulfate and high temperatures(40–70° C). Attachment of enzymes to probes can have a significant effect on the hybridization kinetics and optima of the system. Having the reporter enzyme directly attached to the probe significantly reduces the processing time for blot development after hybridization.

Haptens and cofactors: A wide variety of haptens and cofactors (including biotin, N-acetoxy-N-acetyl-2-aminofluorene, dinitrophenol, fluorescein, digoxigenin, glucose and sulfonate residues) have

been attached to DNA or RNA as distinguishing targets for secondary detection systems. The principal means of detecting these substituents is with antibodies raised against the hapten. Antibodies bound to the hapten-DNA (or -RNA) are detected using conventional immunodetection methods (Figure 10). Direct immunodetection procedures use anti-hapten (primary) antibodies conjugated to a reporter such as an enzyme or fluorophor. Indirect immunodetection procedures use a sandwich method: the unmodified primary antibody is allowed to bind to the hapten-DNA, then the bound primary antibody is detected by a secondary (anti-primary) antibody to which reporter groups are attached.

If biotin is used to label a nucleic acid, the initial binding step commonly uses avidin or streptavidin to detect the labeled probe (Leary *et al.* 1983). These proteins are capable of binding four molecules of biotin and have affinities for biotin <106-fold higher than found in most antibody-hapten binding systems. Reporter groups can be directly conjugated to streptavidin, analogous to direct immunodetection, or the protein can be detected by its ability to bind biotinylated reporter molecules in an indirect method.

Non-radioactive detection: For most blot applications, enzymes are the preferred reporters whether linked directly or indirectly to a probe. Calf intestine alkaline phosphatase and horseradish peroxidase conjugated to secondary antibodies and streptavidin are commercially available. They have well characterized soluble and insoluble chromogenic substrates. When applied to Southern blots, the col-

ored products of the reaction must be insoluble in order to deposit color only at the sites where the reporter enzyme is bound to the filter through the hybridized probe. The chromogenic substrates used are the same as those used for immunocytochemical assays and western blots. Alkaline phosphatase detection is normally 2 to 5 times as sensitive as horseradish peroxidase.

Chemiluminescent methods for detection of both horseradish peroxidase and alkaline phosphatase have been reported recently (Matthews *et al.* 1985; Schaap *et al.* 1987). The removal of the phosphate group from a stabilized dioxetane (Figure 11a) by alkaline phosphatase produces an activated intermediate which decays with the emission of light. In a fluorescein/micelle complex, this system produces light steadily for many hours. For horseradish peroxidase, the breakdown of hydrogen peroxide by the enzyme results in the release of light from luminol (Figure 11b) in the presence of an enhancer such as p-hydroxycinnamic acid. For either system, the emitted light can be detected by a luminometer for dot blots and solution hybridization, or by exposure of high-speed film (x-ray or standard) for Southern or dot blots.

Colloidal gold conjugates of antibodies or strepavidin can be used as a non-enzymatic chromogenic detection system. The gold particles make a reddish-brown deposit on a filter which can be intensified by further treatment to deposit black silver grains on the gold (Tomlinson *et al.* 1988). The sensitivity is similar to that of enzymatic DNA hybrid detection methods.
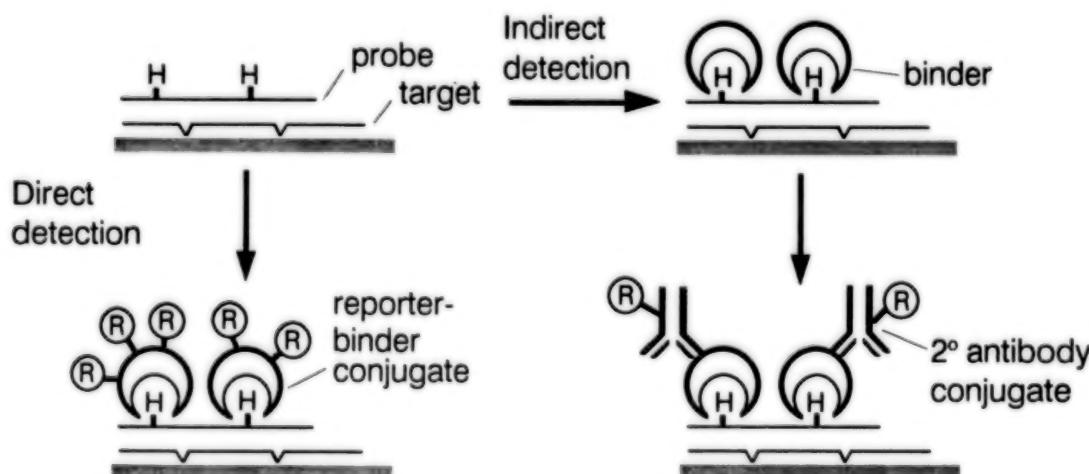
Figure 10. Non-radioactive detection of hapten or biotin-labeled probes. For direct detection of hapten labeled DNA, antibodies to the hapten (or streptavidin if the label is biotin) are conjugated to a reporter molecule (*) such as alkaline phosphatase or fluorescein. This antibody- or streptavidin- reporter conjugate is then allowed to bind to the labeled probe hybridized to the target. Indirect detection uses secondary antibodies conjugated to a reporter to detect the binding of the anti-hapten antibody to a labeled probe.

a. Lumigen™ PPD
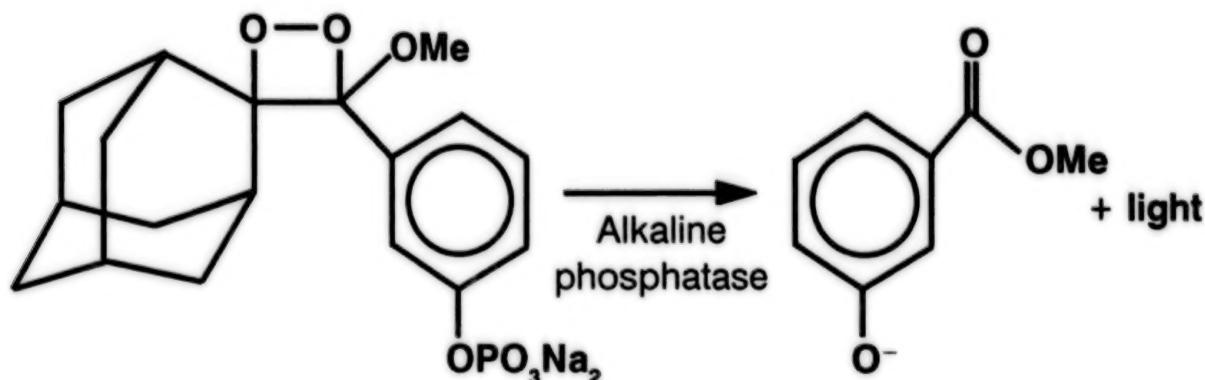
b. Luminol
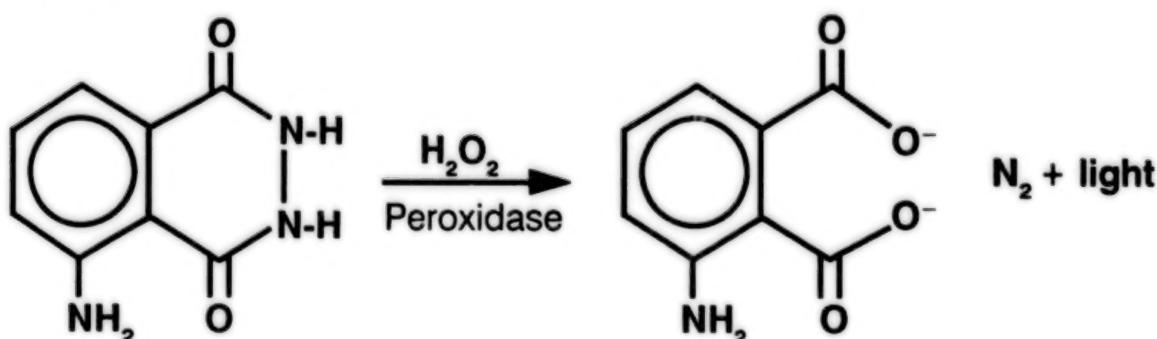
Figure 11. Chemiluminescent precursors. a. Lumigen™ PPD is a stable dioxetane and chemiluminescent substrate for alkaline phosphatase. b. Luminol releases a photon after oxidation by a product of the degradation of hydrogen peroxide by horseradish peroxidase.

CONCLUSIONS

Comparison of P^{32} and non-radioactive detection: Although a great deal of effort has been and continues to be invested in developing sensitive non-radioactive detection methods for nucleic acids, P^{32} autoradiography is currently the method of choice for detection of hybridization in samples where the amount of DNA is limiting (Table 4). The best commercially available colorimetric non-radioactive systems currently yield a signal similar to that of a 4 to 18 hour autoradiograph (Donovan *et al.* 1987). This is satisfactory for laboratory-drawn specimens, but film exposures can reliably be extended to greater than 100 hours to yield the 5- to 20-fold improvement in sensitivity often needed for forensic material. Newer chemiluminescent methods may soon rival P^{32} in sensitivity and offer the same convenient and complete stripability for re-probing.

Table 4. SENSITIVITY OF NUCLEIC ACID DETECTION METHODS

Method	Amount of target DNA (pg)
P^{32} autoradiography	<0.05
P^{32} Geiger-Mueller or scintillation	1–10
Colorimetric	1
Fluorescence	1–100
Chemiluminescence	<0.1

Part of the reason for the greater sensitivity of radioactivity is the linear relationship between the amount of radioisotope incorporated and the signal on the film. The amount of isotope incorporated can be increased by increasing the specific activity of the labeled nucleotide or by using more than one labeled nucleotide. Experiments with biotinylated probes have revealed that the sensitivity of colorimetric detection reaches a plateau beyond approximately 4 biotinylated residues per 100 bases,

presumably due to steric interference between adjoining streptavidin molecules (Gebeyehu *et al.* 1987). Colorimetric detections may also be inhibited by accumulation of precipitated material at the position of the band. Another difficulty with most non-radioactive detection systems is control of non-specific background. The blocking of non-specific binding by reporter proteins to blotting membranes is generally less complete than blocking of binding by unhybridized nucleic acid probes. Further improvements for non-radioactive detection will come primarily in the reporter portion of this system with additional enhancements from changes in membrane blocking techniques.

This paper has discussed only a fraction of the variety of labels, labeling methods and detection methods available for use with nucleic acid probes. Experience has shown that no one system is appropriate for all applications or laboratory situations. This is an area still changing rapidly, even after many years of research. The challenge of practical, routine applications in forensic and medical laboratories is spurring even more activity to simplify, stabilize and improve the sensitivity of nucleic acid probe labeling and detection methodology.

REFERENCES

Al-Hakim, A. H. and Hull, R. (1986). Studies towards the development of chemically synthesized non-radioactive biotinylated nucleic acid hybridization probes, Nucleic Acids Res. 24:9965–9976.

Anonymous (1987). Preparation and use of biotinylated RNA probes, Focus 9(4):10.

Bonner W. M. and Laskey R. A. (1974). A film detection method for tritium-labelled proteins and nucleic acids in polyacrylamide gels, Eur. J. Biochem. 46:83–88.

Connell, C., Fung, S., Heiner, C., Bridgham, J., Chakerian, V., Heron, E., Jones, B., Menchen, S., Mordan, W., Raff, M., Recknor, M., Smith, L., Springer, J., Woo, S. and Hunkapiller, M. (1987). Automated DNA sequence analysis, BioTechniques 5:342–348.

Deen, K. C., Landers, T. A., Berninger, M. (1983). Use of T4 DNA polymerase replacement synthesis for specific labeling of plasmid-cloned inserts, Anal. Biochem. 135:456–465.

Donovan, R. M., Bush, C. E., Peterson, W. R., Parker, L. H., Cohen, S. H., Jordan, G. W., Vanden Brink, K. M. and Goldstein, E. (1987). Comparison of non-radioactive DNA hybridization probes to detect human immunodeficiency virus nucleic acid, Molec. Cellular Probes 1:359–366.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity, Anal. Biochem.132:6–13.

Feinberg, A. P. and Vogelstein, B. (1984). Addendum to "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity," Anal. Biochem. 137:266–267.

Forster, A. C., McInnis, J. L., Skingle, D. C. and Symons, R. H. (1985). Non-radioactive hybridization probes prepared by the chemical labelling of DNA and RNA with a novel reagent, photobiotin, Nucleic Acids Res. 13:745–761.

Gebeyehu, G., Rao, P. Y., Soo Chan, P., Simms, D. A. and Klevan, L. (1987). Novel biotinylated nucleotide analogs for labeling and colorimetric detection of DNA, Nucleic Acids Res. 15:4513–4534.

Gillam, I. C. and Tener, G. M. (1986). N4-(6-aminohexyl)cytidine and deoxycytidine nucleotides can be used to label DNA, Anal. Biochem. 157:199–207.

Green, M., Maniatis, T. and Melton, D. (1983). Human β-globin pre-mRNA synthesized *in vitro* is accurately spliced in Xenopus oocyte nuclei, Cell 32:681–694.

Jablonski, E., Moomaw, E. W., Tullis, R. H. and Ruth, J. L. (1986). Preparation of oligodeoxynucleotide-alkaline phosphatase conjugates and their use as hybridization probes, Nucleic Acids Res. 14:6115–6128.

Keller, G. H., Cumming, C. U., Huang, D.-P., Manak, M. M. and Ting, R. (1988). A chemical method for introducing haptens onto DNA probes, Anal. Biochem. 170:441–450.

Kumar, A., Tchen, P., Roullet, F. and Cohen, J. (1988). Nonradioactive labeling of synthetic oligonucleotide probes with terminal deoxynucleotidyltransferase, Anal. Biochem. 169:376–382.

Leary, J. J., Brigati, D. J. and Ward, D. C. (1983). Rapid and sensitive colorimetric method for visualizing biotin-labeled DNA probes hybridized to DNA or RNA immobilized on nitrocellulose: bio-blots, Proc. Nat. Acad. Sci. USA 80:4045–4049.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 122–127.

Matthews, J. A., Bakti, A., Hynds, C. and Kricka, L. J. (1985). Enhanced chemiluminescent method for the detection of DNA dot-hybridization assays, Anal. Biochem. 151:205–209.

McCracken, S. (1985). Preparation of RNA transcripts using SP6 RNA polymerase, Focus 7(2):5–8.

O'Farrell, P. (1981). Replacement synthesis method of labeling DNA fragments, Focus 3(3):1–3.

Piatyszek, M. A., Jarmolowski, A. and Augustyniak, J. (1988). Iodo-Gen-mediated radioiodination of nucleic acids, Anal. Biochem. 172:356–359.

Prober, J. M., Trainor, G. L., Dam. R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A. and Baumeister, K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides, Science 238:336–341.

Renz, M. and Kurz, C. (1984). A colorimetric method for DNA hybridization, Nucleic Acids Res. 12:3435–3444.

Rigby, P. W. J., Dieckmann, M., Rhodes, C. and Berg, P. (1977). Labeling deoxy-ribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I, J. Mol. Biol. 113:237–251.

Roberts, P. L. (1985). Comparison of fluorographic methods for detecting radioactivity in polyacrylamide gels or on nitrocellulose filters, Anal. Biochem. 147:521–524.

Roychoudhury, R. and Wu, R. (1980). Terminal transferase-catalyzed addition of nucleotides to the 3 termini of DNA. In: Methods in Enzymology, vol 65, part I (Grossman, L. and Moldave, K., eds.). Academic Press, NY, pp. 43–62.

Sanger, F., Nicklen, S. and Coulsen, A.R. (1977). DNA sequencing with chain-terminating inhibitors, Proc. Nat. Acad. Sci. USA 74:5463–5467.

Schaap, A. P., Sandison, M. D. and Handley, R. S. (1987). Chemical and enzymatic triggering of 1,2-dioxetanes. 3: Alkaline phosphatase-catalyzed chemiluminescence from an aryl phosphate-substituted dioxetane, Tetrahedron Lett. 28:1159–1162.

Smith, L. M., Fung, S., Hunkapiller, M. W., Hunkapiller, T. J. and Hood, L. E. (1985). The synthesis of oligonucleotides containing an aliphatic amino group at the 5 terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis, Nucleic Acids Res. 13:2399–2412.

Tchen, P., Fuchs, R. P. P., Sage, E. and Leng, M. (1984). Chemically modified nucleic acids as immunodetectable probes in hybridization experiments, Proc. Nat. Acad. Sci. USA 81:3466–3470.

Tomlinson, S., Lyga, A., Huguenel, E. and Dattagupta, N. (1988). Detection of biotinylated nucleic acid hybrids by antibody-coated gold colloid, Anal. Biochem. 171:217–222.

BLANK PAGE

MULTILOCUS AND SINGLE LOCUS MINISATELLITE DNA PROBES IN FORENSIC MEDICINE

Alec J. Jeffreys, Victoria Wilson, Zilla Wong, Illa Patel, Nicola Royle, Rita Neumann, John A. L. Armour, Robert Kelley, Andrew Collick, Ian Gray and Mark Gibbs

University of Leicester,
Leicester, United Kingdom

A little over four years ago, we described a novel DNA hybridization system which produced highly individual-specific patterns from human DNA, and speculated that these so-called DNA fingerprints might have potential application in many areas of human genetics, including parentage testing and forensic analysis (Jeffreys *et al.* 1985b). Since then, events have moved with great rapidity, not only with the development of an extensive battery of ever more powerful and sensitive DNA typing systems, but also with the rapid diffusion of DNA testing into government and commercial laboratories, and into civil and criminal casework worldwide. The new technology has raised not only scientific issues, but also much wider social, ethical, legal and political issues, concerned for example with civil liberties, the politics of immigration control and the potential undermining of the adversarial nature of the judicial system.

The purpose of this paper is to review the scientific basis of the DNA typing systems we have developed, rather than to explore the non-scientific issues. We shall discuss the basis of human DNA variation, the development of multilocus and single locus minisatellite DNA probes, and more recent advances in the enzymatic amplification of highly variable regions of the human genome.

HUMAN DNA POLYMORPHISMS

Prior to the mid-1970's, biochemical genetics focussed of necessity not on genes but on gene products. While enzyme polymorphisms and cell surface antigen markers have provided a rich source of genetic markers, their use in forensic medicine has been restricted by the limited variability of most markers, by instability of proteins in forensic specimens and by the tissue specificity of many biochemical markers. By the late 1970's, DNA cloning and nucleic acid hybridization technology had developed to the point where direct analysis of sequence polymorphisms in human DNA became possible (Kan and Dozy 1978; Jeffreys 1979). The ubiquity of DNA together with the more recent appreciation that DNA can sometimes show surprising stability, as witnessed by the successful cloning of DNA from a 2400 year old Egyptian mummy (Paabo 1985), made DNA the obvious choice for forensic analysis.

Almost any cloned segment of unique-sequence human DNA can, with enough effort, detect restriction fragment length polymorphisms (RFLPs) in human DNA. Unfortunately, most of this variation results from single base substitutions or microdeletions/insertions which create or destroy restriction endonuclease cleavage sites. The resulting RFLPs are restriction enzyme specific and dimorphic, with only three possible phenotypes per marker; the use of these RFLPs in forensics is therefore, severely limited by their low variability. However, mutational processes other than base substitution also introduce variation into the human genome; these processes include transposition, unequal exchange and gene conversion during recombination between duplicated DNA sequences, and polymerase slippage at replication forks. The latter two processes appear to be responsible for producing localized tandem-repeated minisatellite regions in the human genome. These minisatellites frequently show high levels of allelic variation in the number of repeat units and therefore in the length of the minisatellite. The resulting length variation can be detected using the cloned minisatellite as a hybridization probe on Southern blots of human DNA digested with any restriction enzyme which does not cleave within the array of tandem repeats. In practice, allelic resolution is maximized by digesting human DNA with a restriction endonuclease which recognizes a 4 base pair (bp) sequence and which therefore cleaves frequently in human DNA and removes most of the flanking DNA from a minisatellite or VNTR (variable number tandem repeat) markers provide the most variable, and therefore most individual-specific, multi-allelic loci in the human genome and are ideally suited to forensic analysis.

DNA FINGERPRINTS

The first hypervariable locus in human DNA was discovered by Wyman and White (1980). The

molecular basis of variability remained unknown until DNA sequence analysis of another hypervariable region near the human insulin gene showed it to be tandem-repeated (Bell *et al.* 1982). The overall abundance of hypervariable minisatellites in the human genome appears to be low, as witnessed by large-scale screen of 1680 different recombinants selected at random from a human genomic library; of these clones, only 12 detected variable loci with heterozygosities in excess of 70% (Knowlton *et al.* 1986).

A more selective method for identifying hypervariable loci came from the discovery that a subset of human minisatellites share, with minor variation, a common core sequence embedded in each repeat unit which is presumably involved in the generation of these hypervariable tandem-repeated loci perhaps by serving as a recombination or slippage signal (Jeffreys *et al.* 1985b). Hybridization probes consisting of tandem repeats of this core sequence (polycore probes) detect approximately 1000 loci simultaneously in human DNA. The majority of the minisatellites detected are short and of low variability, and can be removed from Southern blot profiles if human DNA is digested with a restriction endonuclease such as *Hin*fI or *Alu*I which cleaves frequently in human DNA. The larger and more variable minisatellite DNA fragments are resistant to cleavage with such restriction enzymes, and can be visualized with these polycore probes on Southern blots to provide a complex and highly variable DNA fingerprint (Figure 1) (Jeffreys *et al.* 1985b,c) not only from human DNA but also from DNA from a wide variety of other vertebrates (Jeffreys *et al.* 1987b). Two versions of the polycore probes termed 33.6 and 33.15 have been developed. Since these two probes detect essentially non-overlapping sets of human minisatellites (the degree of overlap is less than 20%), they can provide substantially independent DNA fingerprints from the same DNA sample (Jeffreys *et al.* 1985c, 1986).
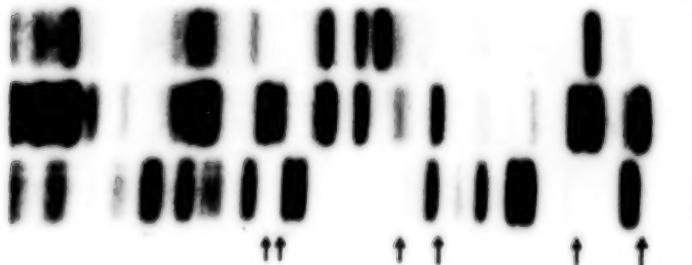
Human DNA fingerprints detected by these polycore probes are highly individual-specific (with the exception of monozygous twins who cannot be distinguished using these or any other known DNA probes) and are simply inherited. Pedigree analysis shows that individual resolved fragments in a human DNA fingerprint behave as heterozygous Mendelian markers and assort independently from parent to offspring. The 50–100 hypervariable loci which contribute to the human DNA fingerprint are widely scattered in the human genome, but are not randomly dispersed in that they tend to localize preferentially to the proterminal regions of human autosomes and frequently show close physical linkage to other hypervariable minisatellites not detected by DNA fingerprint probes 33.6 and 33.15 (Royle *et al.* 1988). The level of allelism or linkage between pairs of DNA fragments detected in a human DNA fingerprint is low, and there is not detectable linkage disequilibrium between any pairs of resolved DNA fragments (Jeffreys *et al.* 1986). This genetic analysis, as well as comparisons of the DNA fingerprints of different individuals in a population, establishes that the different DNA fragments resolved in a human DNA fingerprint act essentially as independent genetic markers. This is however not necessarily true for all vertebrates, and indeed clear evidence for non-independence (linkage disequilibrium) is evident in the DNA fingerprints of mice (Jeffreys *et al.* 1987a).

Two parameters govern the individual specificity of a DNA fingerprint (Jeffreys *et al.* 1985c), namely the number of resolved large (>4 kilobase, Kb) minisatellite DNA fragments detected by probes 33.6 and 33.15 (36 ± 7 (SD) in a survey of 74 unrelated English individuals), and the mean probability that a DNA fragment in one individual will be matched by a band of similar electrophoretic mobility and autoradiographic intensity in a second unrelated individual (0.25, varying from 0.10 for the largest (>10 Kb) DNA fragments to

Figure 1. Resolution of a paternity dispute by DNA fingerprint analysis using probe 33.15. Note that many of the child's paternal DNA fragments (arrowed) are not present in Mr Y, establishing through multiple exclusion that he cannot be the father of this child. This was confirmed by re-testing with probe 33.6 (not shown). Reproduced with permission from Jeffreys *et al.* (1987b).

0.28 for smaller (4–6 Kb) minisatellites; the human DNA fingerprint below 4 Kb consists of a very complex set of poorly resolved DNA fragments and is excluded from probability calculations). Band-sharing between unrelated individuals frequently results from the fortuitous comigration of different minisatellite DNA fragments, rather than representing isoallelism; the band sharing frequency of 0.25 is therefore largely determined by the limitations of gel electrophoretic resolution, and should therefore vary little between different human populations, as has indeed been observed.

The band sharing frequency can be used to provide conservative estimates of the probability of false association in for example forensic analyses and paternity disputes. Since parkers are independent, then for a typical individual, the chance that a second unrelated individual would contain all of the first individual's DNA fingerprint bands detected by probes 33.6 and 33.15 can be estimated at $0.25^{36} = 2 \times 10^{-22}$. This probability estimate is conservative for the following reasons: A) for a full match, the second individual's DNA fingerprint should contain no additional bands not present in the first DNA fingerprint. By imposing this criterion, the probability of false association falls from 2×10^{-22} to 4×10^{-30} (Jeffreys and Morton 1987). B) hertogeneity in band sharing frequencies according to band size will reduce the probability estimate. C) the criteria for deducing a match between two samples, namely correspondence of band number, electrophoretic mobilities and relative hybridization intensities of different bands, are more stringent than the criteria used for estimating the frequency of band sharing. D) matching should extend over the entire DNA fingerprint, including the smaller (<4 Kb) DNA fragments which are excluded from probability calculations. The true probability of false association between unrelated individuals, using both DNA fingerprint probes on undegraded human genomic DNA, is therefore $<< 4 \times 10^{-30}$. Even for first degree relatives, such as sibs, this probability is vanishingly low (approximately 10^{-14}). Therefore, DNA fingerprints provide a means for positively establishing associations, with the proviso that monozygous twins cannot be distinguished.

Similarly, DNA fingerprints provide a single powerful test for positively establishing parentage in for example paternity disputes (Figure 1) (Jeffreys et al. 1985c). If all DNA fingerprint bands in a child are present in the claimed mother and/or father, then maternity and paternity are simultaneously established. In paternity disputes, the significance of association can be determined most simply from p, the number of bands in the child which cannot be attributed to the mother. An estimate of the probability that an unrelated man would by chance contain all of these paternal characters is 0.25^p. In casework, p is typically approximately 16, giving a mean probability of 2×10^{-10} that a falsely accused man would fail to be excluded as the father by DNA fingerprinting. More detailed statistical analysis shows that this estimate is conservative (Hill 1986; J. F. Y. Brookfield, personal communication). It is also conservative for the reasons described above. With this statistical power, DNA fingerprinting can not only positively establish paternity, but can also be used to resolve immigration disputes where all non-parental hypotheses concerning relationship, such as the possibility that the claimed father is a close blood relative of the true father, need to be rejected (Home Office 1988). Also, DNA fingerprinting can be extended to families where one parent is missing, by using undisputed children or other appropriate relatives to reconstruct most or all of the missing parent's DNA fingerprint (Jeffreys et al. 1985a). If a case involves only one parent and a child, then the problem reduces to estimating the degree of band sharing between two individuals. This approach almost always, for authentic claims, allows the hypothesis of no relationship to be rejected, and an estimate of the degree of relationship can be derived. Parentage as such cannot however be proven in cases where only one parent is available, and it is sometimes not possible to distinguish a first-degree from a second-degree relationship such as uncle-nephew. The problems of using DNA fingerprint band sharing to estimate relationships have been discussed in detail by Lynch (1988).

One potential source of uncertainty in parentage testing is minisatellite mutation which could lead to one or more bands in a child being mutant and therefore unattributable to either parent. Mutations altering the number of repeats and therefore the length of a minisatellite have been observed to occur in DNA fingerprints at a mean frequency of 0.006 per offspring band, resulting in 25% of children showing one mutant band in the profiles produced by probes 33.6 and 33.15, about 1% showing two mutations and an estimated 0.05% showing three. Such non-ascribable bands can be readily accommodated in statistical analyses of the probability estimates (J. F. Y. Brookfield, personal communication). In practice, one or two mutant bands do not significantly affect the statistically highly polarized nature of the paternity test results. Three or

more mutant bands can introduce statistical uncertainties which however can usually be resolved by subsequent sequential testing with locus-specific minisatellite probes (see following discussion). In virtually all cases, DNA fingerprinting can therefore give a definitive conclusion concerning parentage.

The DNA fingerprints are highly reproducible, provided that due attention is paid to the quality control of probe preparation, the completeness of restriction endonuclease digestion of human DNA, and gel electrophoresis and Southern blot hybridization procedures. They can derived from a minimum of 0.1–1 µg of high molecular weight DNA. Limited degradation, for example in a forensic specimen, can lead to the preferential loss of the largest DNA fragments in a DNA fingerprint. Such degradation, as well as imbalances in amounts of DNA tested, need to be taken into account in evaluating the similarity of the DNA fingerprint of a forensic specimen with that of a suspect and in calculating probability estimates. Multilocus DNA fingerprints cannot reliably be used to detect mixed samples, for example a semen sample recovered from a victim raped by two men. These probes also cross-react with DNA from a wide range of vertebrates, and the possibility of contamination of a forensic specimen with animal remains needs to be borne in mind.

The multilocus DNA fingerprint system has already been extensively used in casework in the United Kingdom (U. K.). Both for identification and for establishing family relationships, DNA fingerprint evidence has been accepted in civil and criminal courts. In addition, DNA fingerprinting in the U. K. immigration casework has been evaluated, alongside conventional genetic marker systems, in a pilot study of immigrant families carried out jointly by ourselves, Cellmark Diagnostics, the Home Office and the Foreign and Commonwealth Office. This study confirmed that DNA testing was more effective than conventional markers at detecting incorrect family relationships, and interestingly showed that the substantial majority of individuals who had been previously refused entry clearance and who volunteered to take part in the study were in fact related as claimed. Profiling of DNA using DNA fingerprint probes together with locus-specific minisatellite probes (see following discussion) has now been acknowledged as reliable for the routine resolution of immigration disputes in the U. K. (Home Office 1988).

LOCUS-SPECIFIC MINISATELLITE PROBES

The largest minisatellite DNA fragments detected in a human DNA fingerprint contain large numbers of tandem repeat units and are therefore likely to represent the most variable minisatellites. By selectively cloning these large minisatellites, it has proved possible to isolate some of the most variable loci yet described in the human genome (Wong et al. 1986, 1987). To date, 15 hypervariable loci with heterozygosities ranging from 60–99.4% have been isolated by this method (median value approximately 95%) (Figure 2). More extensive collections of VNTR probes have been isolated by screening cosmid libraries by hybridization with oligonucleotides based on known minisatellite repeat units (Nakamura et al. 1987); while most of the loci detected show limited allelic variability, a minority are highly variable with heterozygosities in excess of 90%.
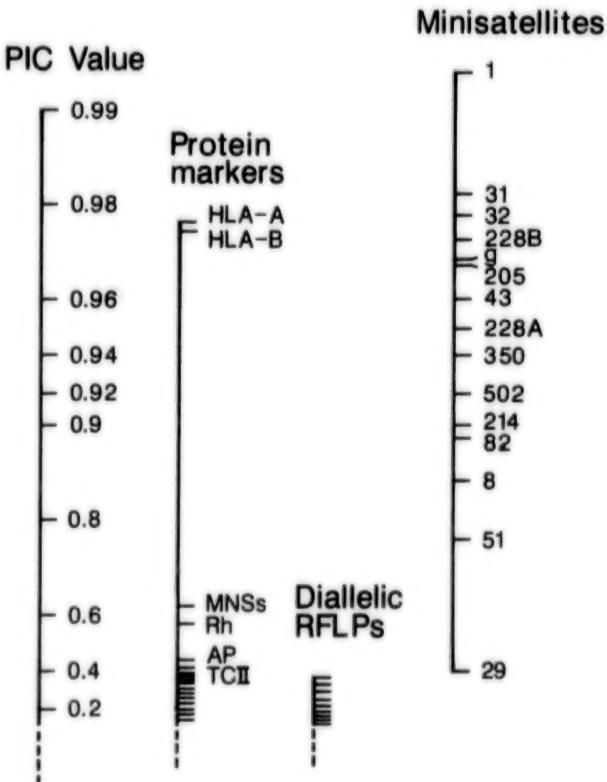
Figure 2. Comparison of the variability of protein markers with diallelic RFLPs and cloned minisatellites. Variability is expressed as the polymorphic information content (PIC) of a marker, which for highly variable loci approximates to the heterozygosity.

Under high stringency hybridization conditions, minisatellites cloned from DNA fingerprints act as locus-specific probes. For the most variable locus, detected by clone MS1, scores of alleles ranging from 1–23 Kb in length can be resolved by

gel electrophoresis. Analysis of the allele frequency distribution in North European populations shows that there are no common alleles at this locus. The mean allele frequency is approximately 0.6%, and the maximum frequency of any allele detected is 3%. These estimates do not reflect true allele frequencies but rather reflect the inability to resolve by gel electrophoresis alleles differing in length by one or a few 9 bp repeat units. For loci with heterozygosities <95%, alleles with a significant population frequency can usually be detected. For example, λMS8 detects a locus with a heterozygosity of 81%, with three common alleles at 4.3, 6.2 and 8.1 Kb (frequency in an English population 36%, 24% and 7%) plus many rare alleles (Wong *et al.* 1987). Localization of these hypervariable minisatellites by somatic cell hybrid analysis and in situ hybridization shows that they are dispersed in the human genome (through preferentially subtelomeric) and therefore act as independent markers (Royle *et al.* 1988).

Extremely variable loci must show a significant level of instability to maintain this level of variability. Indeed, mutation rates to new length alleles at the most variable loci can be directly measured by studying large numbers of human pedigrees (Jeffreys *et al.* 1988a). As expected, the germline mutation rate rises with heterozygosity and can become significant above 95% heterozygosity. The most variable locus, detected by λMS1, has the highest germline mutation rate of 5% per gamete, compared with <1% for other loci tested. Mutations occur with similar frequency in sperm and oocytes, and usually involve the gain or loss of a relatively small number of minisatellite repeat units, resulting in the mutant allele being similar in size to its progenitor.

Minisatellite mutation events are not restricted to the germline but also occur somatically, as shown by the presence of mutant minisatellite alleles in clonal tumour cell populations (Armour *et al.* 1989). In general, loci with the highest germline mutation rate show the greatest level of somatic instability. Even in the most rapidly proliferating tissue (gut epithelium), however, we estimate that <15% of cells contain a mutant allele at the most unstable locus, detected by λMS1. There is evidence that the incidence of mutant cells will be even lower in tissue with a lower rate of cell turnover. Since, in polyclonal tissue, mutant alleles will be heterogeneous in size, then they are most unlikely to be detectable by Southern blot analysis of bulk tissue DNA. In practice, we have seen no evidence for mutational mosaicism in 60 different peripheral blood samples tested with six different locus-specific minisatellites, and only two examples in 475 lymphoblastoid cell lines tested with these probes; in these two cases, a novel allele could be detected in addition to the two progenitor alleles, indicating mosaicism for the new mutant allele in what are probably oligoclonal cell lines. The possibility of mutational mosaicism leading to abnormal three-allele phenotypes is not however hypothetical. We have recently characterized a highly unstable mouse minisatellite with a germline mutation rate of 3% per gamete; in addition, approximately 3% of adult mice show abnormal three-allele phenotypes due to global mosaicism in all adult tissue for a new mutant allele. These somatic mutants apparently arise preferentially very early in development and subsequently spread to all adult tissues, including in some cases the germ-line (R. Kelly, G. Bulfield, A. Collick, M. Gibbs and A. J. Jeffreys, unpublished data). An important aspect of probe validation is therefore extensive population screening to establish that mutational mosaicism is rare.

The hybridization patterns detected by these hypervariable locus-specific human minisatellite probes are very simple and consist of two (occasionally one) hybridizing DNA fragments per individual (Figure 3). Equivalent patterns will obtained using restriction endonuclease which does not cut the minisatellite repeat array. Endonucleases which cut every repeat element destroy the minisatellite and are not suitable for typing. Some endonucleases have been found which cut occasionally at variant minisatellite repeat units to produce from some alleles a haplotype of two or more linked minisatellite DNA fragments instead of one band per allele. Potential variation in the distribution of internal variant sites, producing unusual phenotypes with three or more bands, must be taken into account in forensic analyses. Most hypervariable loci are GC-rich and the most appropriate restriction enzymes are endonucleases such as *Alu*I which cleaves at the sequence AGCT and *Hin*f1 (GANTC), which are unlikely to cleave the minisatellite but will remove most of the flanking DNA to maximize the resolution of alleles. Endonucleases such as *Hae*III (GGCC) which cleave at GC-rich sequences are less appropriate and are likely to generate multi-band haplotypes or destroy a significant proportion of minisatellite/VNTR loci.

Locus-specific minisatellite probes provide a good level of individual specificity, though orders of magnitude less per probe than that of the multilocus DNA fingerprint probes (Wong *et al.* 1987).

important for loci showing both common and rare alleles. Suppose for example that a locus had a common allele at 4.0 Kb ($q = 0.3$) but that alleles at 4.1 Kb are rare ($q = 0.01$). If an allele in the forensic specimen is sized at 4.1 Kb, but with an error likely to exceed 2.5%, then the estimate for q_i will change dramatically. As noted previously, loci with heterozygosities greater than 95% seldom show evidence of common alleles, and an estimate of the probability of false association can be reasonably derived from the mean allele frequency q_m as $2q_m^2$, without any need to measure allele sizes. More conservatively, the mean allele frequency q_m can be arbitrarily set to the frequency of the most common allelic size class q_c, when the probability becomes, $<2q_c^2$. For the most variable locus detected by MS1, $q_m = 0.006$ and $q_c = 0.03$, and the probability of false association becomes 7×10^{-5} and, 1.8×10^{-3} respectively. Sequential testing with a panel of extremely variable locus-specific probes therefore provides a high degree of individual specificity, without the need to size alleles in casework.

Locus-specific minisatellite probes provide a number of advantages for forensic analysis. First, the probes are simpler to use than the multilocus DNA fingerprint probes. Second, they are more sensitive and can type as little as 50 μg human genomic DNA, corresponding to 2 μl blood, <1 μl semen or 20 μl saliva. Approximately 70% of pulled hair roots contain sufficient DNA for typing with these locus-specific probes. Third, they are effective at detecting mixed DNA samples in for example cases of multiple rape, when the presence of four non-victim alleles would indicate that two rapists were involved (or possibly more, depending on allele frequencies). The ability to analyze mixed samples also enables ultravariable probes to be used for mass screening. Suppose, for example, that one had the DNA profile from a forensic specimen and were faced with screening 1000 potential suspects. Thirtytwo pools of 31–32 suspects could initially be screened for a pool containing the two relevant alleles; the relevant individual could then be identified by typing individual members of the positive pool. Thus 1000 individuals could be screened on just two gel electrophoretic runs. A fourth advantage of locus-specific probes is that they are human-specific and rarely cross-hybridize with DNA from other species, unlike the DNA fingerprint probes. Fifth, the patterns obtained are simple and readily interpretable in court. Several of these points are illustrated in Figure 3 which shows the results of an analysis of forensic specimens from

Figure 3. Analysis of samples from the Enderby murder case by Southern blot hybridization with locus-specific minisatellite probes λMS1 and λMS31. Note that the non-victim semen alleles in samples 2, 5 and 6 are indistinguishable, indicating that both victims had been sexually assaulted by the same individual. Only two semen alleles are detected with each prove, establishing that the semen was derived from one man. A suspect charged with the murder of the second victim is clearly not the source of the semen found on both victims and was subsequently released. Further details are given in Wong et al. (1987), Gill and Werrett (1987) and Wambaugh (1989). Reproduced with permission from Jeffreys et al. (1989).

In a forensic analysis of for example a blood stain and a suspect, the significance of a match will be given by $2q_iq_j$, where q_i and q_j are the frequencies of the two alleles detected as determined in an appropriate reference population group. Homozygotes can be similarly analyzed. Since alleles sizes can vary continuously, these q_i and q_j values need to be modified to take into account experimental error in allele length determination. This can be

the Enderby murder case which led to the release of a suspect who had been falsely charged with one of the murders (Wong *et al.* 1987; Gill and Werrett 1987).

Locus-specific minisatellite probes also provide a useful adjunct to paternity testing, to identify the incorrect parent(s) in cases of DNA fingerprint exclusion and to provide additional information in rare cases where DNA fingerprints do not give a definitive conclusion. In cases of non-exclusion with locus-specific minisatellite probes, mutation rates at these loci are largely irrelevant. In cases of exclusion, estimates of mutation rate are needed to estimate the likelihood ratio of correct paternity with mutation versus incorrect paternity. By restricting the analysis to loci with heterozygosities of <95%, this problem is largely avoided (Jeffreys *et al.* 1988a), though at a cost of loss of allelic variability.

AMPLIFICATION OF MINISATELLITES BY THE POLYMERASE CHAIN REACTION

Forensic specimens containing less than 50 ng of high molecular weight human DNA cannot usually be analyzed with multilocus or locus-specific minisatellite probes. We have therefore explored the possibility of enzymatically amplifying specific minisatellite loci from human DNA using the polymerase chain reaction (PCR) (Saiki *et al.* 1988), to permit DNA fingerprint analysis of much smaller amounts of human DNA (Jeffreys *et al.* 1988b).

To amplify minisatellites, oligonucleotide primers from unique sequence DNA flanking a minisatellite are used of drive the amplification of the entire block of minisatellite repeat units. The size of the PCR product is therefore determined by the allelic number of minisatellite repeat units. PCR products are subsequently detected by Southern blot hybridization with a minisatellite probe.

Faithful amplification of minisatellite alleles up to 5–15 Kb long can be achieved. However, fidelity is lost at high numbers of PCR amplification cycles due to out-of-register annealing of minisatellite DNA strands at high concentration of PCR product, resulting in the profile degenerating into a heterodisperse smear of PCR products. For this reason, visualization of PCR products by hybridization is advisable, although with careful control of PCR cycle numbers it is possible to amplify alleles up to 5 Kb long to the point where they can be directly visualized by electrophoresis on an ethidium bromide-stained agarose gel (A. J. Jeffreys, R. Neumann and J. A. L. Armouret, unpublished

data; see also Boerwinkle *et al.* 1989; Horn *et al.* 1989).

Since the PCR reaction is best terminated during the early exponential phase of accumulation of product, there should be little or no interference between different loci being amplified. In practice, at least six different minisatellites can be simultaneously co-amplified from the same DNA sample and the products simultaneously or sequentially detected with appropriate locus-specific minisatellite probes. Simultaneous detection produces a PCR-derived DNA fingerprint (Figure 4) which can be readily obtained from as little as 1 ng genomic DNA and which shows a good degree of individual specificity (the chance of false association between unrelated people is estimated at approximately 2×10^{-5} for the six minisatellite loci so far developed for PCR amplification).
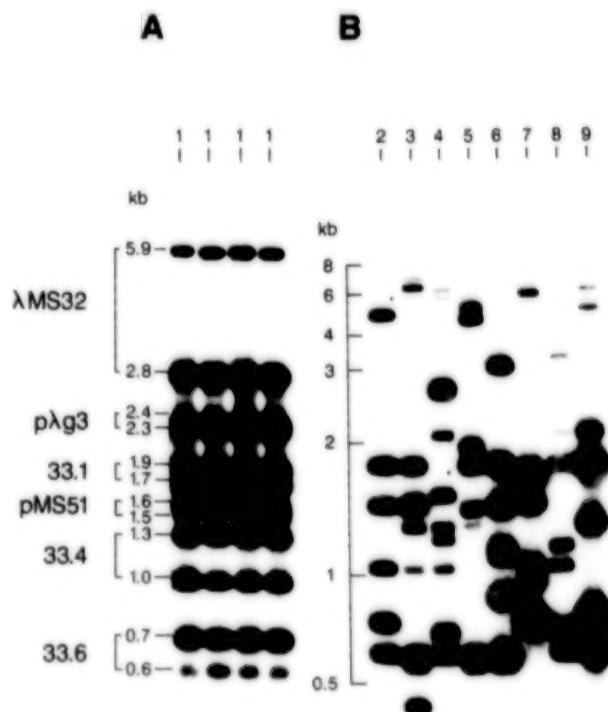
Figure 4. PCR-derived DNA fingerprints obtained by co-amplifying six minisatellite loci simultaneously, followed by simultaneous detection of all six loci with a cocktail of six labelled minisatellite probes. A) repeat analyses of 10 ng genomic DNA from an individual, showing reproducibility of the DNA fingerprint. This individual is heterozygous at all six loci tested. B) PCR DNA fingerprints from 1 ng DNA samples from 8 unrelated individuals (Jeffreys *et al.* 1988b). Reproduced with permission from Jeffreys *et al.* (1989).

In addition, PCR is capable of amplifying single minisatellite molecules and analyzing single human cells (Jeffreys *et al.* 1988b). However, during the large number of amplification cycles needed to produce sufficient PCR product for de-

tection, some loci show, in addition to *bona fide* amplification products, the occasional appearance of spurious products generated during PCR. Another significant problem in our experience is the inadvertent contamination of PCR reactions with minute traces of recombinant DNA or the products of previous PCR reactions. Nevertheless, individual identification based on DNA analysis at the level of one or a few cells is a feasible proposition.

With greatly improved sensitivity provided by PCR, individual-specific DNA typing systems can be extended to the analysis of minute samples such as hair roots, 0.01–0.001 μl blood stains, urine, submicroliter amounts of saliva and microscopic traces of semen. Another advantage is that crude cell lysates can be subjected to PCR (Li *et al.* 1988; Jeffreys *et al.* 1988b), obviating the need for purifying DNA.

MAPPING VARIANT MINISATELLITE REPEATS

Even the most variable human minisatellites seldom show complete uniformity of repeat sequences in all alleles. Instead, variant repeat units are frequently present, often intermingled with the normal repeat units throughout an allele (Jeffreys *et al.* 1985a; Wong *et al.* 1986). Using PCR to amplify minisatellite alleles followed by end-labelling of the PCR product and partial restriction endonuclease digestion of repeat units, it is possible to explore this internal variation as new source of minisatellite variability. Examples of this analysis of internal variation in different amplified alleles from the hypervariable locus detected by minisatellite clone λMS32 are shown in Figure 5. Restriction enzyme *Hinf*I cuts every 29 bp repeat unit to produce a continuous ladder of labelled DNA fragments. The absolute number of repeat units per allele can therefore be established, without reference to molecular weight standards or control DNA samples. In contrast, *Hae*III only cleaves at a variant repeat unit which, on average, accounts for approximately 70% of all repeat units. As a result, the ladder produced by *Hae*III is discontinuous. For any allele, comparison of the continuous *Hinf*I ladder and discontinuous *Hae*III ladder allows each repeat unit to be classified as to whether or not it is cut by *Hae*III, resulting in an unambiguous binary string encoding the number of repeats and location of variant repeats along the allele. In principle, up to 10^{60} different alleles at the locus detected by λMS32 could be distinguished using this system, in contrast to approximately 100 alleles by conventional gel electrophoresis and Southern blot hybrid-

ization. In addition, alleles are distinguished by an unambiguous binary code, rather than by estimated molecular weight (with attendant error margins). It remains to be seen whether such an allele coding system can be technically simplified to the point where it can be routinely used in forensic analysis to establish matches between alleles in a suspect and in a forensic specimen without the need to estimate alleles sizes or to provide side-by-side electrophoretic comparisons of DNA form the suspect and the forensic sample.

REFERENCES

Armour, J. A. L., Patel, I., Thein, S. L., Fey, M. F. and Jeffreys, A. J. (1989). Analysis of somatic mutations at human minisatellite loci in tumors and cell lines, Genomics 4:328–334.

Bell, G. I., Selby, M. J. and Rutter, W. J. (1982). The highly polymorphic region near the insulin gene is composed of simple tandemly repeating sequences, Nature 295:31–35.

Boerwinkle, E., Xiong, W., Fourest, E. and Chan, L. (1989). Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction: application to the apolipoprotein B 3' hypervariable region, Proc. Nat. Acad. Sci. USA 86:212–216.

Gill, P. and Werrett, D. J. (1987). Exclusion of a man charged with murder by DNA fingerprinting, Forensic Science International 35:145–148.

Hill, W. G. (1986). DNA fingerprint analysis in immigration test-cases, Nature 322:290–291.

Home Office (1988). DNA profiling in immigration casework. Report of a pilot trial by the Home Office and Foreign and Commonwealth Office. (Home Office, London).

Horn, G. T., Richards, B. and Klinger, K. W. (1989). Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction, Nucleic Acids Res. 17:21–40.
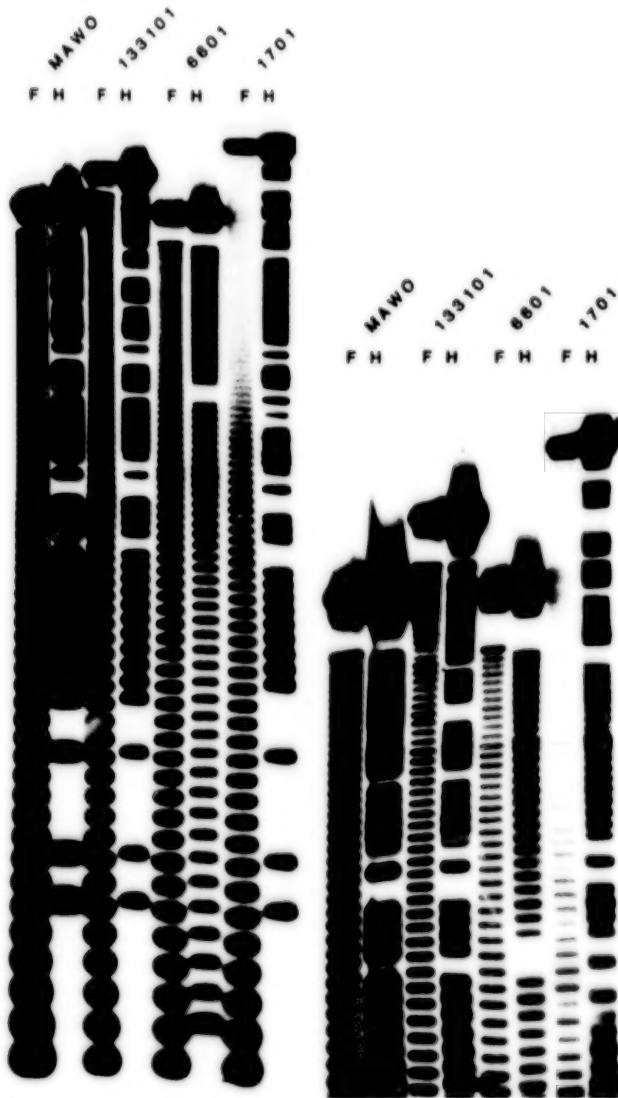
Figure 5. Internal mapping of variant repeat units in alleles of λMS32 amplified by the polymerase chain reaction. Amplified alleles were labelled with 32ₚ at one end of the DNA fragment and partially digested with *Hin*fI (F) or *Hae*III (H), followed by normal (left) or extended (right) gel electrophoresis and autoradiography. *Hin*fI produce a continuous ladder, enabling the number of 29 bp repeat units to be counted in each allele. *Hae*III only cleaves a subset of repeat units to produce an interrupted ladder. Each allele can be converted to a binary string, where 1 = repeat cut by *Hae*III and 0 = repeat not cut. Reading from bottom to top of the gel, the binary string for allele 133101 is thus 00000101000010011111 11111101110010111111 11101111001011111101 11110011110111111111 111010. Note that different length alleles, for example MAWO and 133101, can have closely related internal maps, suggesting common ancestry, and that alleles of similar lengths, such as MAWO and 6601, can have a totally different internal structure.

Jeffreys, A. J. (1979). DNA sequence variants in the G -, A -, and -globin genes in man, Cell 18:1-10.

Jeffreys, A. J., Brookfield, J. F. Y. and Semeonoff, F. R. (1985a). Positive identification of an immigration test-case using human DNA fingerprints, Nature 317:818-819.

Jeffreys, A. J., Wilson, V. and Thein, S. L (1985b). Hypervariable "minisatellite" regions in human DNA, Nature 314:67-73.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985c). Individual-specific "fingerprints" of human DNA, Nature 316:76-79.

Jeffreys, A. J., Wilson, V., Thein, S. L., Weatherall, D. J. and Ponder, B. A. J. (1986). DNA "fingerprints" and segration analysis of multiple markers in human pedigrees, Am. J. Hum. Genet. 39:11-24.

Jeffreys, A. J. and Morton, D. B. (1987). DNA fingerprints of dogs and cats, Animal Genetics 18:1-15.

Jeffreys, A. J., Wilson, V., Kelly, R., Taylor, B. A. and Bulfield, G. (1987a). Mouse DNA "fingerprints": analysis of chromosomes localization and sperm-line stability of hypervariable loci in recombinant inbred strains, Nucleic Acids Res. 15:2823–2836.

Jeffreys, A. J., Wilson, V., Wong, Z., Royle, N., Patel, I., Kelly. R. and Clarkson, R. (1987b). Highly variable minisatellites and DNA fingerprints, Biochem. Soc. Symp. 53:165–180.

Jeffreys, A. J., Royle, N. J., Wilson, V. and Wong, Z. (1988a). Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA, Nature 332:278–281.

Jeffreys, A. J., Wilson, V., Neumann, R. and Keyte, J. (1988b). Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells, Nucleic Acids Res. 16:10953–10971.

Jeffreys, A. J., Wong, Z., Wilson, V., Patel, I., Neumann, R., Royle, N. J. and Armour, J. A. L. (1989). Application of multilocus and single locus minisatellite DNA probes in forensic medicine, Current Communications in Molecular Biology, in press (Cold Spring Harbor).

Kan, Y. W. and Dozy, A. M. (1978). Polymorphism of DNA sequence adjacent to human-globin structural gene: relationship to sickle mutation, Proc. Natl. Acad. Sci. USA, 75:5631–5635.

Knowlton, R. G., Brown, V. A., Braman, J. C., Barker, D., Schumm, J. W., Murray, C., Takvorian, T., Ritz, J. and Donnis-Keller, H. (1986). Use of highly polymorphic DNA probes for genotype analysis following bone marrow transplantation, Blood 68:378–385.

Li, H., Gyllensten, U. B., Cui, X., Saiki, R. K., Erlich, H. A. and Arnheim, N. (1988). Amplifi-cation and analysis of DNA sequences in single human sperm and diploid cells, Nature 335:414–417.

Lynch, M. (1988). Estimation of relatedness by DNA fingerprinting, Mol. Biol. Evol. 5:584–599.

Nakamura, Y., Leppert, M., O'Connel, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping, Science 235:1616–1622.

Paabo, S. (1985). Molecular cloning of Ancient Egyptian mummy DNA, Nature 314:644–645.

Royle, N. J., Clarkson, R. E., Wong, Z. and Jeffreys, A. J. (1988). Clustering of hypervariable mini-satellites in the proterminal regions of human autosomes, Genomics 3:352–360.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, Science 239:487–491.

Wambaugh, J. (1989). The Blooding (Bantam Press, London, NY).

Wong, Z., Wilson, V., Jeffreys, A. J. and Thein, S. L. (1986). Cloning a selected fragment from a human DNA "fingerprint": isolation of an extremely polymorphic minisatellite, Nucleic Acids Res. 14:4605–4616.

Wong, Z., Wilson, V., Patel, I. Povey, S. and Jeffreys, A. J. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA, Ann. Hum. Genet. 51:269–288.

Wyman, A. and White, R. (1980). A highly poly-morphic locus in human DNA, Proc. Nat. Acad. Sci. USA 77:6754–6758.

LABORATORY SAFETY

James L. Mudd

FBI Laboratory
Quantico, Virginia

Forensic scientists have always recognized the need for safe laboratory practices when handling and analyzing the variety of forensic evidence that is routinely encountered. As a result, most forensic laboratories have established guidelines for laboratory safety. Now with the introduction of deoxyribonucleic acid (DNA) typing techniques, such as the restriction fragment length polymorphism (RFLP) method which uses P^{32} labeled DNA probes for detection of the restriction fragments, forensic scientists are faced with new safety concerns associated with the use of radioactive materials.

This article will discuss the various aspects of laboratory safety as applied to the DNA analysis laboratory with particular emphasis on radiation safety. In addition, specific questions of biological and chemical safety as related to the DNA laboratory will be addressed.

RADIATION SAFETY

Exposure to radiation is not confined just to the laboratory (Cheremisinoff *et al.* 1987). There are many natural sources of radiation to which we are exposed. These natural sources of radiation include cosmic radiation, radiation in the air (radon) and sea water. These naturally occurring radioisotopes are eventually transferred to man through the food chain.

In addition to naturally occurring sources of radiation, there are also man made sources of radiation. The largest source of exposure to man made radiation is from the medical community. This exposure comes in the form of diagnostic and dental X-rays, tracers and other types of therapy. There are also industrial uses of radiation (radiography) such as for the inspection of welds.

An atom consist of a nucleus, made up of protons and neutrons, that is surrounded by a cloud of electrons which move in defined orbits or shells around the nucleus. For any element, the number of protons equal the number of electrons.

The atomic structure of a chemical element may exist in more than one configuration. Those elements that have different atomic masses but have the same number of protons in the nucleus (but vary in the number of neutrons) are called isotopes. In some isotopes, the ratio of protons and neutrons in the nucleus is such that the nucleus is unstable. In an attempt to reach a more stable configuration these unstable nuclei will undergo spontaneous nuclear changes by increasing or decreasing the neutron to proton ratio. Isotopes that undergo these spontaneous nuclear changes are called radioisotopes.

The spontaneous changes mentioned above involve the emission of particles and/or electromagnetic radiation from the nucleus by a process of decay or disintegration called radioactivity. There are four types of radioactive decay that will result in the rearrangement of the proton to neutron ratio in the nucleus: alpha, beta, gamma and X-rays (Wang *et al.* 1975). The first two types of decay involve the emission of particles from the nucleus while the latter two methods involve the emission of radiation in the form of electromagnetic waves.

Radioisotopes that undergo alpha decay emit particles that are comprised of two protons and two neutrons. These particles are essentially helium nuclei that have been stripped of their electrons. Emission of an alpha particle, which travels only a few inches in air, results in a decrease in the atomic number by two and a decrease in the atomic mass by four. For example, radon-219 undergoes alpha decay to produce polonium-215, with a decrease in two in the atomic number (86 to 84 respectively) and a decrease of four in atomic mass (219 to 215 respectively).

For those radioisotopes that undergo beta decay, two types of particles can be emitted from the nucleus: beta minus and beta positive particles.

The beta minus particle, called a negatron, is identical to an electron. Decay by the emission of beta minus particles occurs when there is an excess of neutrons in the nucleus and results in the conversion of a neutron to a proton thereby increasing the atomic number by one, but without a change in the atomic mass. For example phosphorus-32 (P^{32}) (atomic number = 15), the radioisotope used in RFLP analysis techniques, undergoes beta minus decay to sulfur-32 (atomic number = 16).

The emission of a beta positive particle, called a positron, occurs when there is an excess of protons in the nucleus. Ejection of these positively

charged particles occurs when a proton decays to a neutron and results in a decrease by one in the atomic number with no change in the atomic mass. For example, nitrogen-13 (atomic number = 7) undergoes beta + decay to carbon-13 (atomic number = 6).

Unlike alpha particles, beta particles can travel quite far in air, approximately 12 feet per MeV (million electron volt) of energy. Therefore, the high energy beta particle (1.71 MeV) ejected by phosphorus-32 can travel more than 20 feet in the air.

Decay by gamma and X-rays are similar in that they both involve the release of excess energy in the form of electromagnetic radiation. Gamma rays are produced by transformations that occur within the nucleus and are usually emitted immediately following the emission of alpha or beta particles. Emission of gamma rays do not produce a change in the number of protons or neutrons. On the other hand, X-rays are produced, in general, from transformations involving the orbital electrons, such as electron capture.

The standard unit of radioactivity is called the curie (Ci) and was originally defined as the number of disintegrations occurring in one gram of pure radium per second and is equal to 3.7×10^{10} disintegrations per second or 2.2×10^{12} disintegrations per minute. The Ci unit of radioactivity refers to the actual number of disintegrations that occur in a sample and not to the number of disintegrations that are detected by a radiation counter.

The rate that a radionuclide decays is indicated by its half-life and is proportional to the number of radioactive nuclei present in a given sample. The half-life of a sample is the time required for the activity of a radionuclide to decrease by one-half. Therefore, with a half-life of approximately 14 days for phosphorus-32, an initial 1.0 Ci concentration of phosphorus-32 would decay to 0.' Ci after one half-life (14 days) and to 0.25 Ci after two half-lives (28 days).

The unit of radiation exposure is the roentgen (R). It is the unit for gamma or X-ray exposure and is a measure of energy dissipation in air. One roentgen will deposit 87.6 ergs of energy per gram of dry air. It cannot be applied to particulate radiation or to tissue absorption.

The rad (radiation absorbed dose) is used to define the amount of ionizing radiation absorbed by tissues due to an exposure of radiation. One rad will deposit 100 ergs of energy per gram of material.

Since the amount of radiation damage to tissues depends upon the type of radiation, its energy and the tissue involved, consideration to these must be given when comparing radiation exposure. The unit used to equalize the biological effects from equal absorbed doses of different types of radiation is the rem (roentgen equivalent man) which is equal to the dose in rads times the quality factor of the radiation adsorbed. Currently, the permissible exposure limits to ionizing radiation are given in rems and are shown in Table 1. These limits do not apply to pregnant women and minors whose exposure limits are limited to 10% of the normal dose limits (0.125 rem per quarter/0.5 rem per year whole body exposure).

Table 1. MAXIMUM PERMISSIBLE OCCUPATIONAL EXPOSURE (REM) AS DEFINED IN TITLE 10 CODE OF FEDERAL REGULATIONS, PART 20.101.

REMS PER CALENDAR QUARTER	
Whole body, head and trunk, active blood forming organs, lens of eyes or gonads	1.25
Hands and forearms, feet and ankles	18.75
Skin of whole body	7.5

When considering the biological effects of exposure to ionizing radiation it is necessary to consider the type of ionizing radiation and the tissues exposed. The types of possible health effects that can be caused by exposure to ionizing radiation are cancer, hereditary effects, dose related effects and prenatal exposure (Federal Register 1987; Nuclear Regulatory Commission 1987). The severity of the effect from exposure are classified as acute for more than 100 rem exposure in less than 24 hours; delayed effects caused by a single acute exposure or a continuing low level exposure (chronic); and genetic effects that impact on the offspring of the exposed individual. Generally, the most sensitive cells are those that multiply rapidly: cells in cancers, skin, fetuses and embryos. The least sensitive cells are those which are highly specialized: nerve, muscle and bone cells.

Because of the potential risks associated with prenatal radiation exposure, female employees working in a radionuclide laboratory should be so advised. This may be accomplished by providing female employees a copy of Nuclear Regulatory Commission Guide 8.13, titled "Instruction Concerning Prenatal Radiation Exposure" (Nuclear Regulatory Commission 1987).

As an external hazard, exposure to P^{32} is normally not serious as it involves only the most superficial layers of the skin. Nevertheless, when dealing with a high energy beta emitter, such as P^{32}, exposure to the hands and other parts of the body should be monitored to ensure that the exposure does not exceed the permissible limits.

Beta emitters pose a significant hazard if taken internally (ingested) (Wang et al. 1975). An isotope like P^{32} may be readily incorporated into the cells of living tissue. If the ingestion leads to a local concentration in the cellular components, the radiation damage may be quite extensive.

While P^{32} does not pose a significant external hazard to the skin, there is a direct correlation between ionizing radiation exposure to the lens of the eye and cataract formation (Radiation Safety Associates 1984). Although the risks of cataract formation is significantly reduced if the exposure occurs over a long period of time, steps should be taken to prevent unnecessary exposure of the lens of the eyes to ionizing radiation. Due to the density and thickness of the tissue preceding the lens of the eye, radionuclides that emit beta particles of less than 0.8 MeV will be stopped before impacting the lens. However, the beta particles emitted by P^{32}, with an energy of 1.71 MeV, will pass through these tissues to the lens unless additional shielding was provided for the eyes. One way to reduce the exposure to the lens of the eyes to high energy beta particles is to work behind suitable beta shielding, such as plexiglass or glass, and to wear safety glasses while in the laboratory.

It should be noted that based on federal guidelines (Title 10, Code of Federal Regulations, Part 20.101) any dose delivered to the lens of the eye, such as P^{32}, must be counted as a whole body dose. Therefore, the permissible exposure limits for P^{32} are those limits set for whole body exposure (1.25 rem per quarter/5.0 rem per year). All personnel who work with radioactivity should use some type of personal monitor such as a body badge (either film or thermoluminescent dosimeter-TLD type) or ring badge (TLD type).

In the laboratory, exposure of an individual to potentially hazardous materials can occur by means of inhalation, ingestion, contact with the eyes or skin, and injection. Therefore, the best way to minimize internal and external hazards, prevent contamination of personnel and equipment and to comply with existing Nuclear Regulatory Commission rules is through the use of safe laboratory practices:

1. Eating, drinking, smoking, applying cosmetics, and storage of food are prohibited in radionuclide work areas. No personal belongings should be stored in the radionuclide work area.

2. Mouth pipetting of any material is forbidden.

3. Protective clothing and gloves must be worn when handling any radioactive material.

4. All radioactive waste and contaminated equipment or glassware must be labeled.

5. Safety glasses should be worn when working with any radioactive material.

6. Lab coats or gloves must not be worn outside the laboratory.

7. Hands must be washed and monitored before leaving the laboratory.

8. Equipment and work areas should be periodically checked for contamination.

Additional information concerning NRC guidelines can be found in Title 10, Code of Federal Regulations (CFR), parts 19 and 20.

BIOLOGICAL SAFETY

As noted previously, safe laboratory practices are an important tool in preventing unnecessary exposure to hazardous material. When dealing with the analysis of biological specimens, there is the potential for exposure to a number of infectious diseases such as hepatitis B, acquired immune deficiency syndrome (AIDS), tetanus, and tuberculosis (Bond 1987; Bigbee 1987). Of these, the greatest risk comes from hepatitis and AIDS when dealing with liquid blood or body fluid specimens in which the AIDS virus is viable for up to 15 days at room temperature or indefinitely if frozen (Resnick et al. 1986). In dried stains at room temperature, however, the AIDS virus is viable for only 72 hours (Centers for Disease Control 1987). Also, common disinfectants will kill the virus within one minute (Bigbee et al. 1989).

Most of the radiation safety practices noted above would also be covered in the safety rules that apply to biological and chemical laboratories. In particular, mouth pipetting should never be allowed in any laboratory. Also, it is important to avoid the production of aerosols when handling either liquid or dried biological specimens and to routinely decontaminate work surfaces. Where possible, a biological safety cabinet should be used when working with liquid biological specimens. Additional personal protection, beyond the use of lab coats, gloves and safety glasses, can also be provided through the use of immune prophylaxis, such as a vaccine against hepatitis.

In general, The Centers for Disease Control recommendations for preventing hepatitis transmission (Bond 1987), would also apply to the other biological hazards, including AIDS, that one might likely encounter in a forensic laboratory.

CHEMICAL SAFETY

In addition to the hazards posed from biological specimens and exposure to ionizing radiation, the individuals who perform DNA RFLP analysis are also routinely exposed to a number of hazardous chemicals such as phenol, chloroform, formamide and ethidium bromide (National Research Council 1981).

Both phenol and chloroform, which are used in the purification of extracted DNA are readily adsorbed through the skin. Phenol, which is corrosive to the eyes and skin, may cause severe burns in concentrated solutions. Additionally, phenol has shown a mutagenic/tumorigenic effect in animals. Therefore, this reagent should be used in a hood and gloves should be worn when handling this reagent.

Chloroform, which is listed as a carcinogen by the National Toxicology Program, is highly volatile, and exposure may cause dizziness, nausea, and headaches as well as creating a cancer risk. Consequently, like phenol, this reagent should be handled with gloves in a fume hood.

The use of formamide in the assay poses a particular risk to pregnant women. This reagent is classified as an embryotoxin which causes adverse effects on the fetus (National Research Council 1981). These effects include death of the fertilized egg, the embryo, or the fetus (embryolethality), malformations (teratologic effects), retarded growth, and postnatal functional deficits. Because the period of greatest susceptibility to embryotoxins is in the first 8 to 12 weeks (during which time some women may not know they are pregnant), women of child bearing age should only handle this substance in a hood and should take precautions to avoid skin contact with the liquid because of the ease with which it passes through the skin. It is recommended that during pregnancy, women should avoid all contact and exposure to this reagent.

Ethidium bromide, which inserts itself (intercalates) between the base pairs of the DNA double helix, is used on the test gels to visualize DNA migration following electrophoresis (Watson *et al.* 1987). This reagent is classified as a mutagen and may alter genetic material. Accordingly, individuals should use gloves and lab coats to avoid contact with eyes, skin, and clothing. It should only be used in a hood.

SUMMARY

In the DNA laboratory the forensic scientist routinely comes in contact with a variety of potentially hazardous radionuclides, biological specimens and chemical reagents. In order to reduce the risks that may be associated with the use of these materials, laboratory workers must adopt and follow rules for safe laboratory performance. The major routes by which exposure can occur in the laboratory are: inhalation, ingestion, skin or eye contact or injection. By adopting such measures of personal protection as the use of gloves, lab coats, eye protection; proper disposal of contaminated material and decontamination of work surfaces; the use of automatic pipetting equipment; and prohibiting the eating, drinking, smoking, application of cosmetics or the storage of food in the laboratory work area, workers would significantly reduce the risk of exposure to radiation, biological or chemical hazards.

In addition, female workers should be acutely aware of the additional hazards during pregnancy associated with the use of radioactive materials and chemicals, such as formamide and ethidium bromide, in RFLP analysis techniques.

REFERENCES

Bigbee, P. D. (1987). Collecting and handling evidence infected with human disease-causing organisms, FBI Law Enforcement Bulletin 56:7:1–5.

Bigbee, P. D., Sarin, P. S., Humphreys, J. C., Eubanks, W. G., Sun, D., Hocken, D. G., Thornton, A. T., Adams, D. E. and Simic, M. G. (1989). Inactivation of human immunodeficiency virus (HIV) by ionizing radiation in body fluids and serological evidence, J. Forensic Sci. 34:1303–1310.

Bond, W. W. (1987). Safety in the Forensic Immunology Laboratory. In: The Proceedings of the International Symposium on Forensic Immunology. U. S. Government Printing Office, Washington, D.C.

Centers for Disease Control. (1987). Recommendations for Prevention of HIV Transmission in Health Care Settings. MMWR, Vol. 36, Supplement No. 2S, p. 4S.

Cheremisinoff, N. P., Cheremisinoff, P. N. and Teresinski, M. F. (1987). Industrial Radiation Hazards Deskbook. Technomic Publishing Co., Lancaster, PA.

Federal Register. (1987). Federal Radiation Protection Guidance of Occupational Exposure. Vol. 52, No. 17, Tuesday, January 27, pp. 2822–2834.

Nuclear Regulatory Commission (1981). Prudent Practices for Handling Hazardous Chemicals in Laboratories. Prepared by the National Research Council's Committee on Hazardous Substances in the Laboratory, National Academy Press, Washington, D.C.

Nuclear Regulatory Commission. (1987). Instruction Concerning Prenatal Radiation Exposure. U. S. Nuclear Regulatory Commission, Regulatory Guide 8.13 (Task OP 031–4), Revision 2, 1987.

Radiation Safety Associates. (1984). Fundamentals of Radiation Protection (1984). Radiation Safety Associates, Inc., Hebron, CT.

Resnick, L., Veren, K., Salahuddin, S., Tondreau, S. and Markham, P. (1986). Stability and inactivation of HTLV–III/LAV under clinical and laboratory environments, JAMA 225:1887–1891.

Wang, C. H., Willis, D. L. and Loveland, W. D. (1975). Radiotracer Methodology in the Biological, Environmental, and Physical Sciences. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A. and Weiner, A. M. (1987). Molecular Biology of the Gene. 4th ed. Vol. 1. Benjamin/Cummings Publishing Co., Menlo Park, CA.

BLANK PAGE

THE USE OF THE POLYMERASE CHAIN REACTION FOR GENETIC TYPING IN FORENSIC SAMPLES

Henry A. Erlich and Russell Higuchi

Department of Human Genetics
Cetus Corporation
Emeryville, California

Cecillia von Beroldingen

Department of Public Health
University of California, Berkeley

Edward Blake

Forensic Science Associates
Richmond, California

The ability to identify genetic polymorphism in DNA extracted from biological evidence samples has revolutionized forensic analysis. The use of Southern blotting and radioactive hybridization probes has allowed the detection of restriction fragment length polymorphisms (RFLP's) that derive from variable number tandem repeat (VNTR) regions of the human genome. This approach requires that DNA be relatively undegraded and present in sufficient quantity (>50 ng). However, many forensic samples cannot be analyzed by this technique because the amount of biological material is too limited and/or the DNA it contains is too degraded. The recent development of polymerase chain reaction (PCR) (Mullis and Faloona 1987; Saiki *et al*. 1985; 1988a; Erlich *et al*. 1988), a method for specific DNA amplification, has overcome these limitations. As a result, it has now become possible to determine the DNA type of a single hair (Higuchi *et al*. 1988), a single cell, (Li *et al*. 1988), and of highly degraded DNA samples (Bugawan *et al*. 1988a) in a simple and rapid test. In this paper, we will describe the use of PCR in the genetic typing of forensic samples.

POLYMERASE CHAIN REACTION

The capacity to amplify specific segments of DNA, made by the PCR enable a new approach to genetic typing. The PCR is an in vitro method for the enzymatic synthesis of specific DNA sequences, using two oligonucleotide primers that hybridize to opposite strands and flank the region of interest in the target DNA. These oligonucleotides anneal to the template DNA such that their 3' OH ends, which are the growing ends of a newly synthesized strand, face each other. When the new strand that results from the extension by DNA polymerase of one primer extends past the other primer site, that strand becomes a new template (Figure 1). Because the primer extension products synthesized in one cycle can serve as a template in the next, the number of target DNA copies doubles approximately at every cycle. Thus, a repetitive series of cycles involving template denaturation, primer annealing, and the extension of the annealed primers by DNA polymerase results in the exponential accumulation of a specific fragment whose termini are defined by the 5' ends of the primers. Twenty cycles of PCR yields about a million fold (2^{20}) amplification. Each cycle consists of (1) a heating step to separate the two DNA strands, (2) a primer annealing step, and (3) a primer extension step at the optimum temperature for DNA polymerase.

The DNA polymerases are enzymes that can synthesize a new strand of DNA from deoxyribonucleotide triphosphates using a template DNA strand and a primer with a 3' OH end to initiate the copying process. Initially, PCR was carried out with the Klenow fragment of *E. coli* DNA polymerase 1. This enzyme has a temperature optimum of 37 °C and is inactivated at the high temperatures required for strand separation. The thermolability of the Klenow fragment required fresh enzyme to be added during every cycle of the PCR. However, the introduction of a thermostable DNA polymerase (*Taq* polymerase) in the first cycle of PCR eliminated the need for any further enzyme additions (Saiki *et al*. 1988a). The use of this enzyme, isolated from the thermophilic bacterium, *Thermos acquaticus*, not only simplified the procedure (now automated) but also significantly increased the
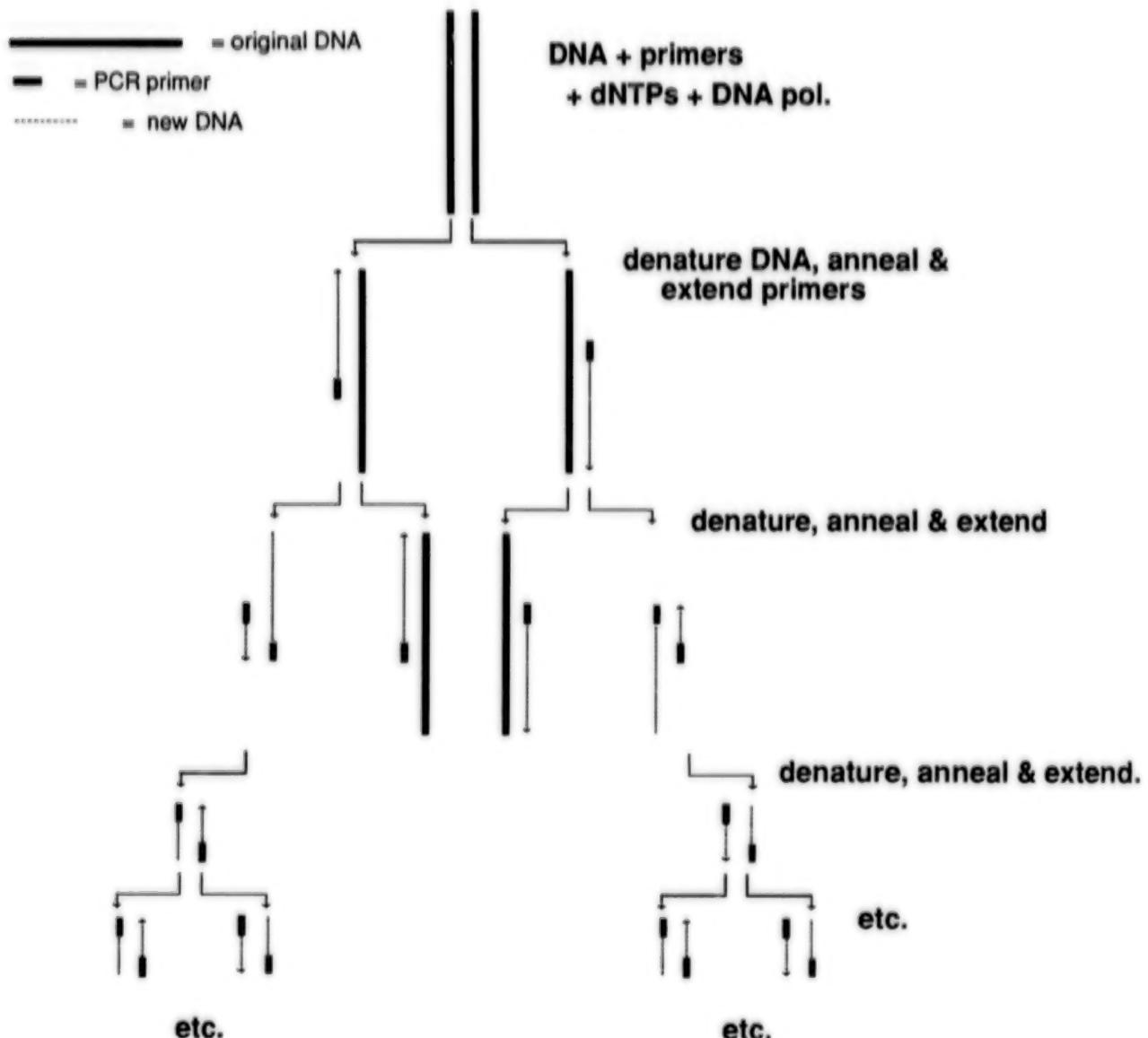
Figure 1. Four Cycles of the Polymerase Chain Reaction (PCR)

specificity (amplification of specific target vs. non-target sequence) and the overall yield of PCR (Saiki *et al*. 1988a). Furthermore, much longer fragments (up to 5 kilobase with reasonable yields) can be amplified using the *Taq* polymerase. The upper limit for PCR amplified DNA fragments from genomic DNA using the Klenow enzyme is about 400 base pairs (Saiki *et al*. 1988a).

GENETIC TYPING

At present, only a few marker systems based on PCR are sufficiently well characterized for use in forensic DNA analysis. As a result, the power of discrimination of these systems is not as high as those employed in RFLP analysis. However, there are many polymorphic regions of human DNA that can be exploited by PCR methods so that the level of discrimination possible should soon improve.

Polymorphisms of DNA can be divided into two categories: sequence polymorphisms, such as occur within the genes of the highly polymorphic HLA complex (Trowsdale *et al*. 1988) and the D-loop region of mitochondrial DNA (Aquadro and Greenberg 1983); and length polymorphism, exemplified by the VNTR (Nakamura *et al*. 1987). In both cases PCR may be used to accumulate DNA fragments containing the polymorphic region. Thereafter, different detection strategies are utilized to distinguish allelic variants.

The most straightforward, albeit laborious, method of identifying sequence polymorphism is to determine the DNA sequence of the amplified product using standard sequencing procedures. In cases where only a single sequence variant is present, for example, in mitochondrial DNA, direct sequencing of the PCR product would be unambiguous (Wrischnik *et al.* 1987). In heterozygous individuals containing two allelic variants, directly sequencing the PCR produce would yield a mixture of two sequences (Gyllensten and Erlich 1988). Identifying the two different sequences without DNA cloning can be problematic, especially if the two alleles differ in sequence at multiple positions.

Sequence polymorphism in the amplified DNA can be detected more simply and rapidly using allele-specific oligonucleotide (ASO) probes (Saiki *et al.* 1986). Under the appropriate conditions, ASO probes will hybridize only to those sequences to which they are perfectly matched. In order to design ASO probes, the DNA sequence of all the commonly occurring allelic variants must first be determined. Once this has been established, ASO probes can be designed to detect these alleles by differential hybridization.

For the analysis of amplified loci with many allelic variants, we have found that the use of non-radioactively-labeled ASO or sequence-specific oligonucleotide (SSO) (In some cases, the hybridization of an oligonucleotide probe does not uniquely specify an allele because the specific sequence is present in more than one allelic variant. In the absence of allele-specific sequences, a given allele is identified as a pattern of SSO probe binding.) probes is the most general and convenient approach. The use of these probes in a dot blot procedure is a powerful and rapid genetic testing method that has been used for the diagnosis of sickle-cell anemia (Bugawan *et al.* 1988a) and β-thalassemia (Saiki *et al.* 1988b) as well as for HLA genotyping (Saiki *et al.* 1986; Scharf *et al.* 1988; Bugawan *et al.* 1988b). This approach involves the PCR amplification of a specific region (that is, the polymorphic second exon of the DQα locus) and the subsequent immobilization of the amplified DNA to replicate filters (nylon membranes). Each filter is then hybridized with an SSO probe labeled either with P^{32}, biotin, or with an enzyme such as horseradish peroxidase. If labeled with an enzyme, the bound probe is detected by the enzymatic conversion of a colorless soluble substrate to a colored precipitate.

For a locus with *n* alleles, each amplified sample must be immobilized on *n* membranes and each membrane hybridized to one of *n* labeled probes. Thus the procedural complexity of this approach is a function of the number of oligonucleotide probes required for complete genetic analysis. Cetus Corporation has recently developed a reverse dot-blot procedure in which the oligonucleotide probe is immobilized on a membrane and hybridized to a labeled PCR product (Saiki *et al.* 1989). In this method, a panel of oligonucleotide probes is UV-crosslined to a nylon membrane through a tail of polythymidine attached to the oligo probe. The PCR product, labeled during amplification by using biotinylated primers, is then hybridized to the immobilized array of oligonucleotide probes. The presence of the specifically bound PCR product is detected using a streptavidin-horseradish peroxidase conjugate (Figure 2). Since each membrane contains all of the allele-specific probes, only one membrane strip need be used to analyze each forensic sample at a particular genetic locus. In fact, genetic typings for multiple loci can be carried out on the same membrane strip, since multiple loci can be simultaneously amplified. This method, described in detail by Saiki *et al.* 1989, has been applied to detection of a variety of β-thalassemia mutations and HLA–DQα alleles. Both the dot-blot and the reverse dot-blot method represent rapid and precise approaches for typing HLA class II gene polymorphism.

These oligonucleotide typing methods provide a qualitative (+ or −) allele definition based on specific probe hybridization unlike the definition based on the electrophoretic mobility of a DNA fragment used in RFLP typing. This means that two samples need not be analyzed at the same time for comparisons to be made and results interpreted. Furthermore, standardization if not a problem and genotype data-bases are easier to compile and use.

Length variation among alleles at a given VNTR locus can be readily detected by size fractionation of PCR products on an analytical gel (Boerwinkle *et al.* 1989; Jeffreys *et al.* 1988; Horn *et al.* 1989). These products are made using primers that flank the repeat region of the locus. As long as the specificity of the reaction is high and no spurious products are generated, the size of amplified fragments should reflect the number of tandem repeats contained within each allele. If the yield and/or specificity of amplification is low, to identify the specific products derived from the VNTR locus, the gel can be blotted and hybridized with a probe specific for the tandem repeat sequence.

The first and most well-developed system for PCR analysis of forensic samples is the DQα
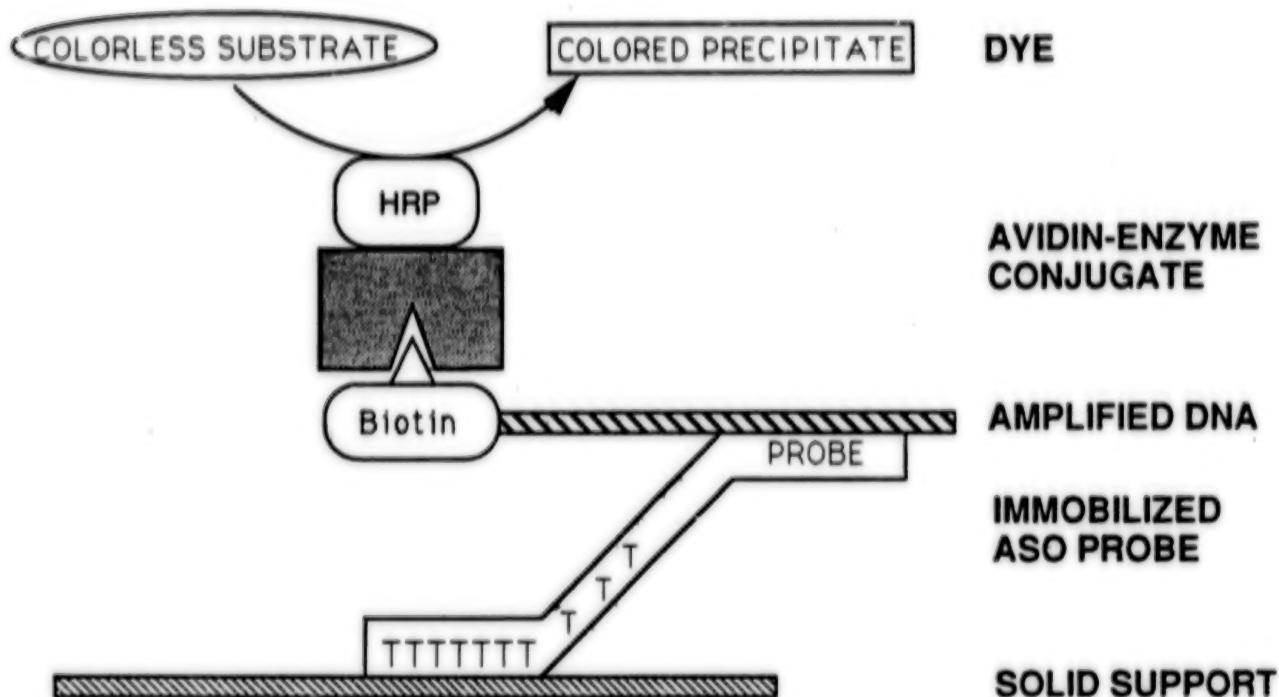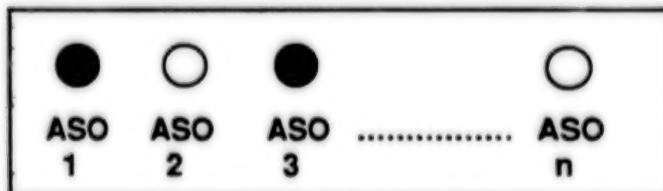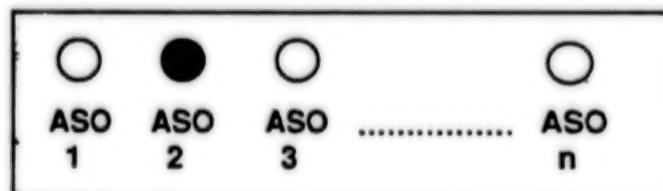
Figure 2. Format II. Immobilized ASO Probes

system. The DQα gene is a class II HLA gene which contains a region of allelic variation within the second exon. Using primers that anneal to conserved sequences flanking this region, a 242 base pair amplification product is obtained. Thus far, eight alleles have been demonstrated by cloning and sequencing amplification products from various serologically defined cell lines (Gyllensten and Erlich 1988; Horn et al. 1989) (Figure 3). Oligonucleotide probes have been designed which define the six most common alleles. Four of the probes hybridize to the same sequence variant region of the amplified product and distinguish the A1, A2, A3, and A4 allelic types. The A1 allelic type may be further subtyped as A1.1, A1.2, or A1.3 by using another set of SSO probes that bind to a dif-

ferent sequence variant region upstream from the first (Figure 3). DQα typing of DNA can be simply and rapidly accomplished by dot-blot hybridization in either of the two formats previously described.

Allele and genotype frequency data have been accumulated for Caucasian, Black, and Asian populations and analysis of the data indicates that these populations are all in Hardy-Weinberg equilibrium for the DQα polymorphism (Erlich, H. A. and coworkers, unpublished). There is no excess of observed homozygotes relative to expected frequencies, consistent with the absence or low frequency of null or blank alleles that might fail to type with the SSO probes. The six alleles distinguishable by this set of probes define twenty-one genotypes with frequencies ranging from less than 0.005 to 0.15. The discriminating power (DP) (Horn *et al.* 1989) of the DQα typing system, that is, the probability of distinguishing between two individuals chosen at random from these populations together, is 0.93. This compares favorably with the discriminating power of traditional genetic marker systems such as ABO blood groups (DP=0.60) and the isoenzyme PGM (DP=0.76) (Sensabaugh 1982). Although far from approaching individual identification, the DQα system has proved useful in excluding or including suspects in criminal cases where conventional genetic typing has failed or where insufficient DNA was available for RFLP analysis.

Unlike the polymorphic VNTR markers, which have a high and measurable mutation rate (Armour *et al.* 1989), HLA sequence polymorphisms are very stable over evolutionary time periods. The analysis by PCR of polymorphic DQα sequences from a variety of primate species suggests that most of the allelic diversity in the contemporary human population was present in the ancestral species that gave rise to the human, chimpanzee, and gorilla lineages (Gyllensten and Erlich 1989). The ability of the same pair of DQα primers to amplify a specific fragment from all individuals from all the primate species tested suggests that the sequences complementary to the primers are highly conserved. Consequently, the existence of human DQα alleles that would fail to amplify with these primers seems highly unlikely.

Another typing system utilizing polymorphisms in the HLA–D region is based upon sequence variation at the DPβ locus. So far, 21 allelic variants have been identified within the highly polymorphic second exon of the DPβ gene based on PCR-derived sequence information (Bugawan *et al.* 1988b; 1989; Erlich and Bugawan 1989). The variant sequences are largely localized to six regions and can be distinguished by a panel of 15 sequence-specific oligonucleotide (SSO) probes. Typing of amplified DPβ sequences is carried out in a dot-blot format, just as with DQα, but the interpretation of the probe hybridization pattern is
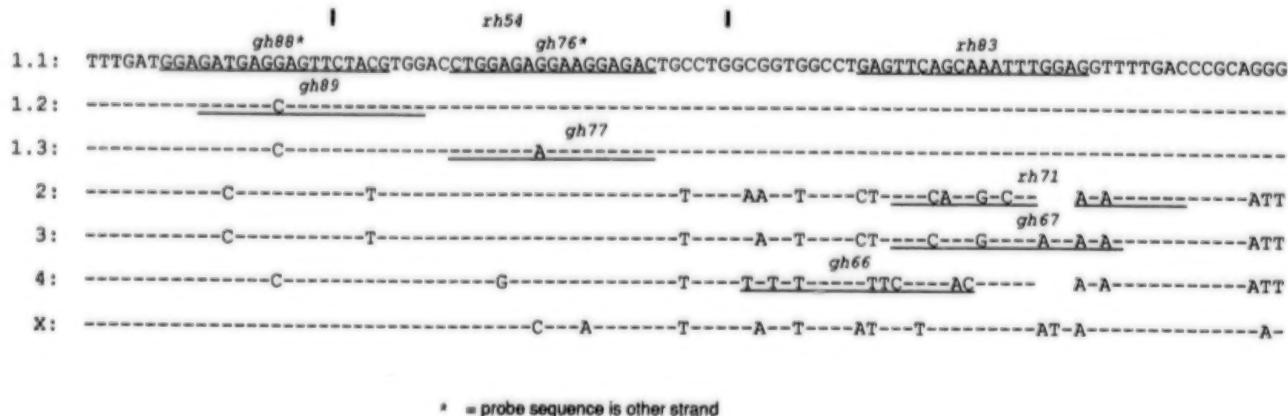
Figure 3. Probes for DQ Alleles

probe	specificity	length	sequence, 5' to 3'
RH54	all	33-MER	CTACGTGGACCTGGAGAGGAAGGAGACTGCCTG
RH83	1	19-MER	GAGTTCAGCAAATTTGGAG
GH88	1.1	19-MER	CGTAGAACTCCTCATCTCC
GH76	all but 1.3	17-MER	GTCTCCTTCCTCTCCAG
GH89	1.2, 1.3, 4	19-MER	GATGAGCAGTTCTACGTGG
GH77	1.3	17-MER	CTGGAGAAGAAGGAGAC
RH71	2	21-MER	TTCCACAGACTTAGATTTGAC
GH67	3	19-MER	TTCCGCAGATTTAGAAGAT
GH66	4	19-MER	TGTTTGCCTGTTCTCAGAC

more complex. This is because few of the probes are specific for a particular allele; rather, different combinations of probes bind to different alleles. Nevertheless, DPβ can be used as a highly informative marker system for individual identification. Furthermore, preliminary results indicate that, in general, DPβ alleles are in linkage equilibrium with DQα alleles (Bugawan, A. and Begovich, H. E., personal communication), thereby allowing the power of discrimination to increase multiplicatively when these systems are sued in combination.

Another potential marker system is the D-loop region of human mitochondrial DNA. Sequence analysis of this region indicates a high degree of variability (Aquadro and Greenberg 1983). Sequence variation may be detected either by direct sequencing or by dot-blot hybridization with a panel of SSO probes. An advantage of mitochondrial DNA is the 10–10,000 fold higher copy number (depending on the tissue) of mitochondrial sequences compared to single copy genomic sequences. When dealing with limited amounts of DNA, the likelihood of amplifying a mitochondrial sequence is correspondingly higher.

Recently, Jeffreys et al. (1988) have reported the amplification of VNTR alleles up to about 10 kilobase in length. At least six different VNTR loci may be co-amplified from nanogram amounts of human DNA and are simultaneously detected by Southern blot hybridization of the products with a battery of locus-specific probes. The resulting DNA fingerprint has a degree of individual specificity approaching 10^{-5}. Multilocus amplification was also obtained from single human cells, although artificial alleles were sometimes generated.

Several factors affect the efficiency and accuracy of VNTR amplifications. In general, there is an inverse relationship between the size of the VNTR allele and the efficiency of amplification: the shorter the allele, the more efficient the amplification. Since the length of VNTR alleles at a particular locus can be quite disparate, the longest alleles may not be detected. Increasing extension time can increase the product yield for longer alleles. However, there still remains a significant molar imbalance between short and long allele products. Increasing cycle number can exacerbate this imbalance and can also generate spurious amplification products such as single-stranded products, partially extended products, or artifactual alleles arising from out-of-register annealing of VNTR templates. This latter phenomenon becomes evident where there is a high concentration of PCR product in the reaction.

Theoretically, the most faithful and efficient amplification of VNTR alleles should be obtained at loci where the alleles are not widely disparate in length and where the amplified with comparable efficiencies. For certain VNTR loci, it has proven possible to detect allelic variants on an ethidium-bromide-stained agarose gel (Boerwinkle et al. 1989; Jeffreys et al. 1988; Horn et al. 1989). For other loci, Southern blot hybridization may still be necessary to detect VNTR alleles. The use of non-radioactive probes, however, would simplify such an analysis, making it adaptable to the typical forensic laboratory (Scharf, S. J. and Erlich, H. A., unpublished data).

Finally, there are a number of dimorphic loci with allele frequencies approaching 50% that amenable to PCR analysis in a dot-blot format, such as the low density lipoprotein receptor (LDLr) gene (Hobbs et al. 1987). Simultaneous amplification and typing of many of these loci, each on different chromosomes, can provide a typing system with good discrimination power (Saiki, R. K. and Erlich, H. A., personal communication). Chamberlain et al. (1988) have amplified up to nine different regions at once. A Y chromosome-specific marker can also be amplified by PCR (Kojan et al. 1987). The male specific product can be directly visualized on an ethidium bromide-stained gel or be analyzed in a dot-blot format.

In summary, a number of highly informative PCR-based genetic typing systems are under development which should make PCR analysis more useful in individual identification. Nevertheless, given that many evidence samples are not suitable for RFLP analysis, we feel that the existing PCR systems will already be of much use. Furthermore, the detection methods used in PCR analysis are rapid, uncomplicated, and do not require the use of radioisotopes. Typing methods involving hybridization with sequence specific oligonucleotide probes also confer the advantage that alleles are defined qualitatively in a plus or minus manner. Therefore, the assembly and use of population data bases is simplified.

AUTOMATION

There are already a number of commercially available automated programmable thermal cycling devices for performing PCR. The production of $>10^6$ copies of a specific sequence can be carried out in an automated two hour reaction following the addition of the sample to a tube containing the reaction reagents. The reverse dot blot method of oligonucleotide probe typing is not currently auto-

mated but is a very simple, rapid procedure for analyzing the amplified products with minimal hands-on time. The complete process including sample preparation analysis, and interpretation takes less than one day.

Instruments that automate some of the steps involved in nucleotide sequence analysis are available commercially. The most rapid and convenient way to carry out an automated sequencing reaction is to prepare the sequencing template by PCR amplification. The misincorporation rate of the *Taq* polymerase in PCR is sufficiently low (1×10^{-10} nucleotides/cycle) (Saiki *et al.* 1988; Tindall and Kunkel 1988), that it does not create problems for sequence analysis of the PCR amplification products. Any individual misincorporated nucleotide will not be detected in the analysis of the population of amplified molecules. Only if the PCR products are cloned and individual clones are sequenced can the misincorporation be detected. In fact, this procedure was used to estimate the misincorporation frequency in PCR (Saiki *et al.* 1988). Recent experiments suggest that varying some of the PCR reaction parameters can lower the error rate (Fucharoen *et al.* 1989). In any case, the misincorporation does not pose a problem for oligonucleotide probe typing because it is the population of amplified molecules that is being analyzed.

CASEWORK

The PCR and HLA–DQα oligonucleotide typing system has been used for forensic casework since summer, 1986. The first introduction of DNA evidence into a U.S. criminal course case was Pennsylvania vs. Pestinikis (December 1986) where various autopsy samples were analyzed to see whether or not they were derived from the same individual (Bugawan *et al.* 1988a). HLA–DQα typing has been carried out, as of August 1989, on 91 different cases including over 1,000 different evidence samples. These samples have included both sperm and vaginal epithelial cells from semen stains, as well as blood stains, single hairs, and tissue and bone fragments. Thus far, in 9 of these cases, the evidence has been introduced into the court proceedings.

The results of the analysis carried out on a semen stain in a sexual assault case are shown in Figure 4. In this well-known case, Gary Dotson was accused of committing a rape in 1977 and convicted in 1979. The alleged victim, Cathleen Webb, recanted in 1985 saying she had fabricated the rape charge vs. Dotson, but her revised story was not accepted by the court. The reference samples from the alleged assailant (G. Dotson) type as a DQA1.1/4 and those from the alleged victim (C. Webb) as DQA1.1/DQA3. The sperm fraction from several different semen stain (now 11 years old) was typed as DQA2/DQA3, consistent with
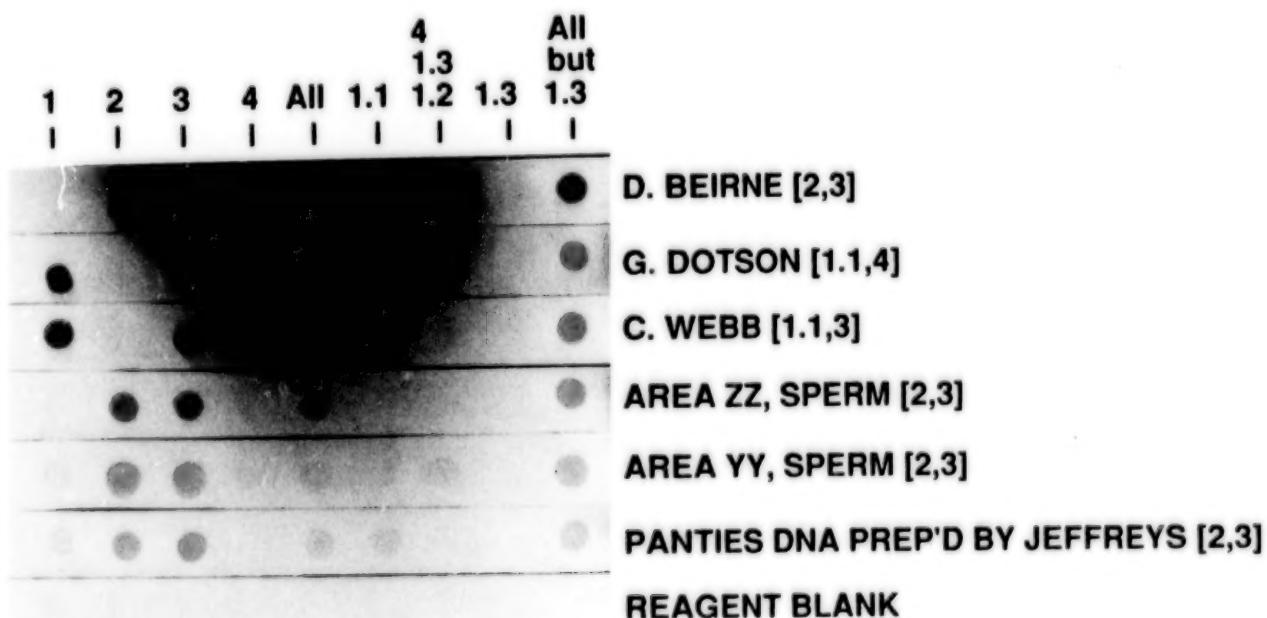
Figure 4. HLA-DQ Typing on Dotson Case Samples

99

the boyfriend's (D. Beirne) DQA type and with the Webb recantation. Also included is a DNA sample prepared from the evidence semen stain which yielded no RFLP pattern when analyzed by Jeffreys because the DNA was too degraded. This sample also typed as DQA2/DQA3 but shows some contamination with Webb's vaginal epithelial cells which were not separated from the sperm in this preparation.

In general, the comparison of the genetic tying of vaginal (or rectal) epithelial cells recovered from semen stains with the type of the sexual assault victim offers an estimate of the reliability of the PCR/HLA-DQα typing method on forensic samples. In over 50 cases, the genotype of the epithelial cells matched the victim reference sample; in some cases, the genetic typing reflected residual DNA from sperm but was consistent with DQα type the victim's reference sample. In one case, the epithelial cell genotype did not match the alleged reference sample. In this case, as was confirmed by a variety of genetic tests, it turned out that the reference sample initially provided was not from the victim. The correct reference sample proved to be an HLA–DQα match with the typing of the vaginal epithelial cells. This DQα genotype was also compatible with the genotype of the victim's parents.

Given the capacity of PCR to synthesize millions of DNA copies, contamination of the sample reaction with either products of a previous reaction (product carryover) or with material from an exogenous source is a potential problem for genetic typing of forensic samples. Careful laboratory procedure, pre-aliquoting reagents, and taking care that the reactions are set up in a space physically separate from where the reactions are set up in a space physically separate from where the reactions are analyzed are all precautions that minimize the risk. A detailed discussion of this issue is presented by Higuchi and Kwok 1989. A sample that has been contaminated can often be identified by a genotyping result with more than two alleles. The use of PCR makes possible repeated testing of the same forensic sample (either in the same lab or in different labs), thereby reducing the chance that a sporadic error (including sample mix-up) would go undetected.

SUMMARY

The ability of PCR to amplify polymorphic regions from limited or degraded DNA samples has expanded the potential of genetic typing to virtually all forensic samples. The availability of simple, rapid, and nonradioactive genetic typing for a variety of genetic markers promises to have a major impact on forensic analysis.

REFERENCES

Aquadro, C.F. and Greenberg, B.D. (1983). Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals, Genetics 103:287–312.

Armour, J.A., Patel, I., Thein, S.L., Fey, M.F. and Jeffreys, A.J. (1989). Analysis of somatic mutations at human minisatellite loci in tumors and cell lines, Genomics 4:328–334.

Boerwinkle, E., Xiong, W., Fourest, E. and Chan, L. (1989). Rapid typing of tandemly repeated hypervariable loci by the polymerase chain reaction: application to the apolipoprotein B 3' hypervariable region, Proc. Nat. Acad. Sci. USA 86:212–216.

Bugawan, T.L., Saiki, R.K., Levenson, C.H., Watson, R.M. and Erlich, H.A. (1988a). The use of nonradioactive oligonucleotide probes to analyze enzymatically amplified DNA for prenatal diagnosis and forensic HLA typing, Biotechnology 6:943–947.

Bugawan, T.L., Horn, G.T., Long, C., Mickelson, E., Hansen, J., Ferrara, G.B., Angelini, G. and Erlich, H.A. (1988b). Analysis of HLA–DP allelic sequence polymorphism using the in vitro enzymatic DNA amplification of DPα and DPβ loci, J. Immunol. 141:4024–4030.

Bugawan, T.L., Angelini, G., Larrick, J., Auricchio, S., Ferrara, G.B. and Erlich, H.A. (1989). A combination of a particular HLA–DPβ allele and an HLA–DQ heterodimer confers susceptibility to coeliac disease, Nature 339:470–472.

Chamberlain, J.S., Gibbs, R.A., Ranier, J.E., Nguyen, P.N. and Caskey, L.T. (1988). Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification, Nucl. Acids Res. 16:11141–11156.

Erlich, H.A., Gelfand, D. and Saiki, R.K. (1988). Specific DNA amplification, Nature 331:461–462.

Erlich, H.A. and Bugawan, T.L. (1989). HLA class II gene polymorphism: DNA typing, evolution, and relationship to disease susceptibility,

In: PCR Technology: Principles and Applications of DNA Amplification. (Erlich, H. ed.), Stockton Press, NY, pp. 193–208.

Fucharoen, S., Fucharoen, G., Fucharoen, P. and Fukumaki, Y. (1989). A novel ochre mutation in the β-Thalassemia gnen of a Thai, J. Biol. Chem. 264:7780–7783.

Gyllensten, U.B. and Erlich, H.A. (1988). Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA–DQ locus, Proc. Natl. Acad. Sci. USA 85:7652–7656.

Gyllensten, U.B. and Erlich, H.A. (1989). Ancient roots for polymorphism at the HLA–DQ locus in primates, Proc. Natl, Acad. Sci. USA, in press.

Higuchi, R., von Beroldingen, C.H., Sensabaugh, G.F. and Erlich, H.A. (1988). DNA typing from single hairs, Nature 332:543–546.

Higuchi, R. and Kowok, S. (1989). Avoiding false positives with PCR, Nature 339:237–238.

Hobbs, H.H., Esser, V. and Russel, D.W. (1987). Avall polymorphism in the human LDL receptor gene, Nucl. Acids Res. 15:379.

Horn, G.T., Richards, B. and Klinger, K.W. (1989). Amplification of a highly polymorphis VNTR segment by the polymerase chain reaction, Nucl. Acids Res. 17:2140.

Horn, G.T., Bugawan, T.L., Long, C. and Erlich, H.A. (1988). Allelic sequence variation of the HLA–DQ loci: Relationship to serology and insulin-dependent diabetes susceptibility, Proc. Natl. Acad. Sci. USA 85:6012.

Jeffreys, A.J., Wilson, V., Newman, R. and Keyle, J. (1988). Amplification of human minisatellites by the polymerase chain reaction: towards DNA fingerprinting of single cells, Nucl. Acids Res. 16:10953–10971.

Kogan, S.C., Doherty, M. and Gitschler, J. (1987). An improved method for prenatal diagnosis of genetic diseases by analysis of amplified DNA sequences. Application to hemophilia A, N. Engl. J. Med. 317, 985–990.

Li, H., Gyllensten, U.B., Cui, X., Saiki, R.K., Erlich, H.A. and Arnheim, N. (1988). Amplification and analysis of DNA sequences in single human sperm and diploid cells, Nature 355:414–417.

Mullis, K.B. and Faloona, F. (1987). Specific synthesis of DNA in vitro via a polymerase catalyzed chain reaction, Meth. Enzymol. 155:335–350.

Nakamura, Y., Leppert, M., O'Connell, P., Wolf, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping, Science 235:1616–1622.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H. A. and Arnheim, N. (1985). Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia, Science 230:1350–1354.

Saiki, R.K., Bugawan, T.L., Horn, G.T., Mullis, K.B. and Erlich, H.A. (1986). Analysis of enzymatically amplified β-globin and HLA–DQα DNA with allele-specific oligonucleotide probes, Nature 324:1350–1354.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S., Higuchi, R. H., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988a). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, Science 239:487–494.

Saiki, R. K., Chang, C. A., Levenson, C. H., Warren, T. C., Boehm, C. D., Kazazlan, Jr., H. H. and Erlich, H. A. (1988b). Diagnosis of sickle cell anemia and -thalassemia with enzymatically amplified DNA and non-radioactive allele-specific oligonucleotide probes, New Engl. J. Med. 319:537–541.

Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. (1989). Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes, Proc. Nal. Acad. Sci. USA 86:6230–6234.

Scharf, S. J., Friedmann, A., Brautbar, C., Szafer, F., Steinman, L., Horn, G., Gyllensten, U. and Erlich, H. A. (1988). HLA class II allelic variation and susceptibility to Pemphigus vulguris, Proc. Natl. Acad. Sci. USA 85:3504–3508.

Sensabaugh, G. F. (1982). Biochemical Markers of Individuality, In: Handbook of Forensic Sciences, Saferstein, R. (ed), Prentice Hall, Englewood Cliffs, NJ. pp. 338–415.

Tindall, K. R. and Kunkel, T. A. (1988). Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase, Biochem. 27:6008–6013.

Trowsdale, J., Young, J. A. T., Kelly, A. P., Austin, P. J., Carson, S., Meunier, H., So, A., Erlich, H. A., Spielman, R., Bodmer, J. and Bodmer, W. F. (1985). Structure, sequence and polymorphism in the HLA-D region, Immunol. Rev. 85:5–43.

Wrischnik, L. A., Higuchi R. H., Stoneking, M., Erlich, H. A., Arnheim, N. and Wilson, A. C. (1987). Length mutations in human mitochondrial DNA; direct sequencing of enzymatically amplified DNA, Nuc. Acids Res. 15:529–535.

BLANK PAGE

DNA SEQUENCING

Catherine T. Comey

FBI Laboratory
Quantico, Virginia

Deoxyribonucleic Acid (DNA) sequencing is the determination of the order of nucleotides in DNA strands. It is a way of detecting polymorphisms, or differences in DNA sequences, at their most fundamental level. In this paper, methods of DNA sequencing, sequence analysis and possible forensic applications will be discussed.

DNA SEQUENCING METHODS

Many methods of DNA sequencing have been described (see Hindley 1983 for a review; Hyman 1988). However, two methods have been used to determine virtually all of the published sequence data. These are the chemical degradation method (Maxam and Gilbert 1977, 1980), and the dideoxynucleotide chain termination method (Sanger *et al.* 1977).

Maxam and Gilbert Sequencing

The chemical degradation method of DNA sequencing involves the generation of fragments of DNA, the ends of which correspond to particular
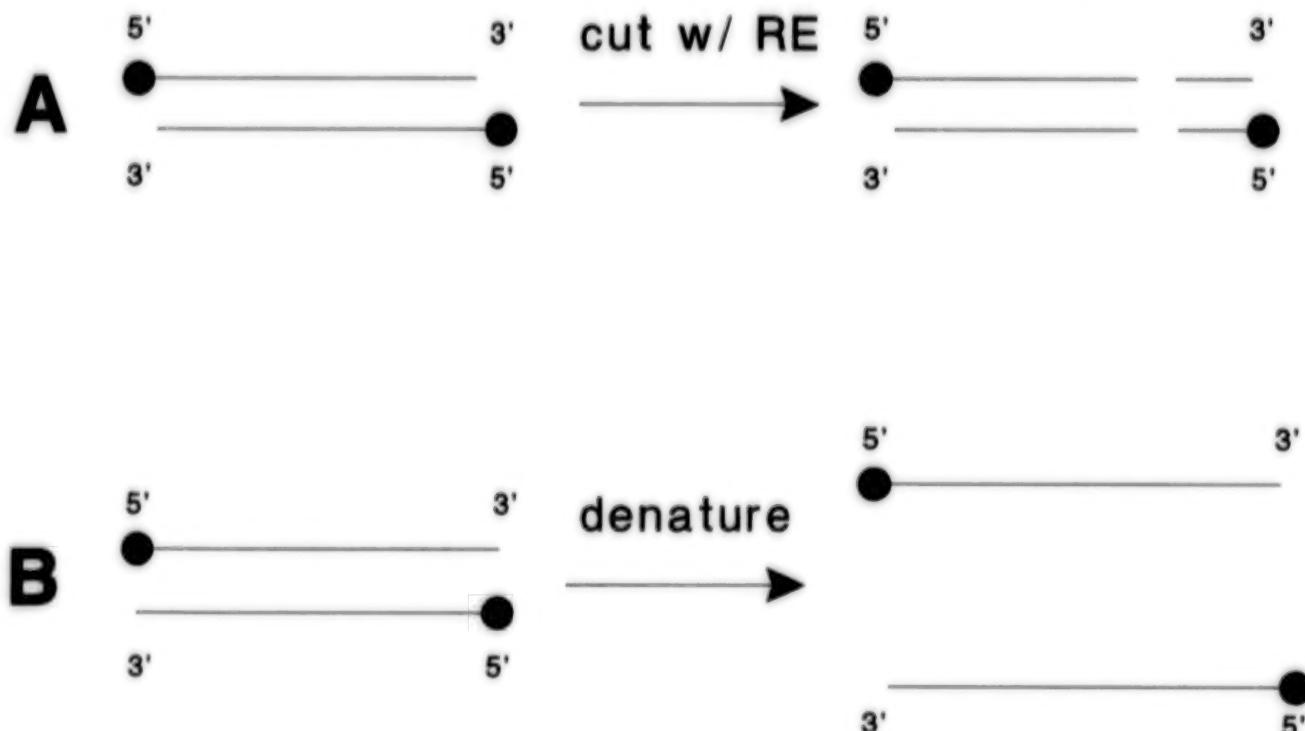
Figure 1. Isolation of a DNA fragment with one radioactive label. Panel A: A double-stranded DNA molecule which has been labelled on its 5′ ends is shown. After cleavage by a restriction enzyme (RE), the two resulting fragments can be separated by gel electrophoresis and isolated, yielding two double-stranded fragments, each having one radioactive label. Panel B: A double-stranded DNA molecule which has been labelled on its 5′ ends is shown. After denaturation of the molecule with NaOH, the strands can be separated by acrylamide gel electrophoresis and isolated (although the strands are the same length, their molecular weights will differ due to differing base composition and cause their separation by electrophoresis). Two single-stranded DNA molecules are obtained, each with a 5′ label.

nucleotides in the piece of DNA being sequenced (Maxam and Gilbert 1977, 1980). So that the fragments can be detected by autoradiography, the initial step in this sequencing method entails the end-labelling of the piece of DNA being sequenced. This can be accomplished by one of several methods (Hindley 1983). These methods label either the 5' or 3' ends of DNA molecules; a double-stranded molecule will have both strands labelled. However, for sequencing, it is required that only one strand be labelled. Thus, after the DNA fragment is end-labelled, either the double-stranded molecule must be made single-stranded and one of the single strands isolated, or the double-stranded molecule must be cut with a restriction enzyme and one of the fragments isolated (Figure 1).
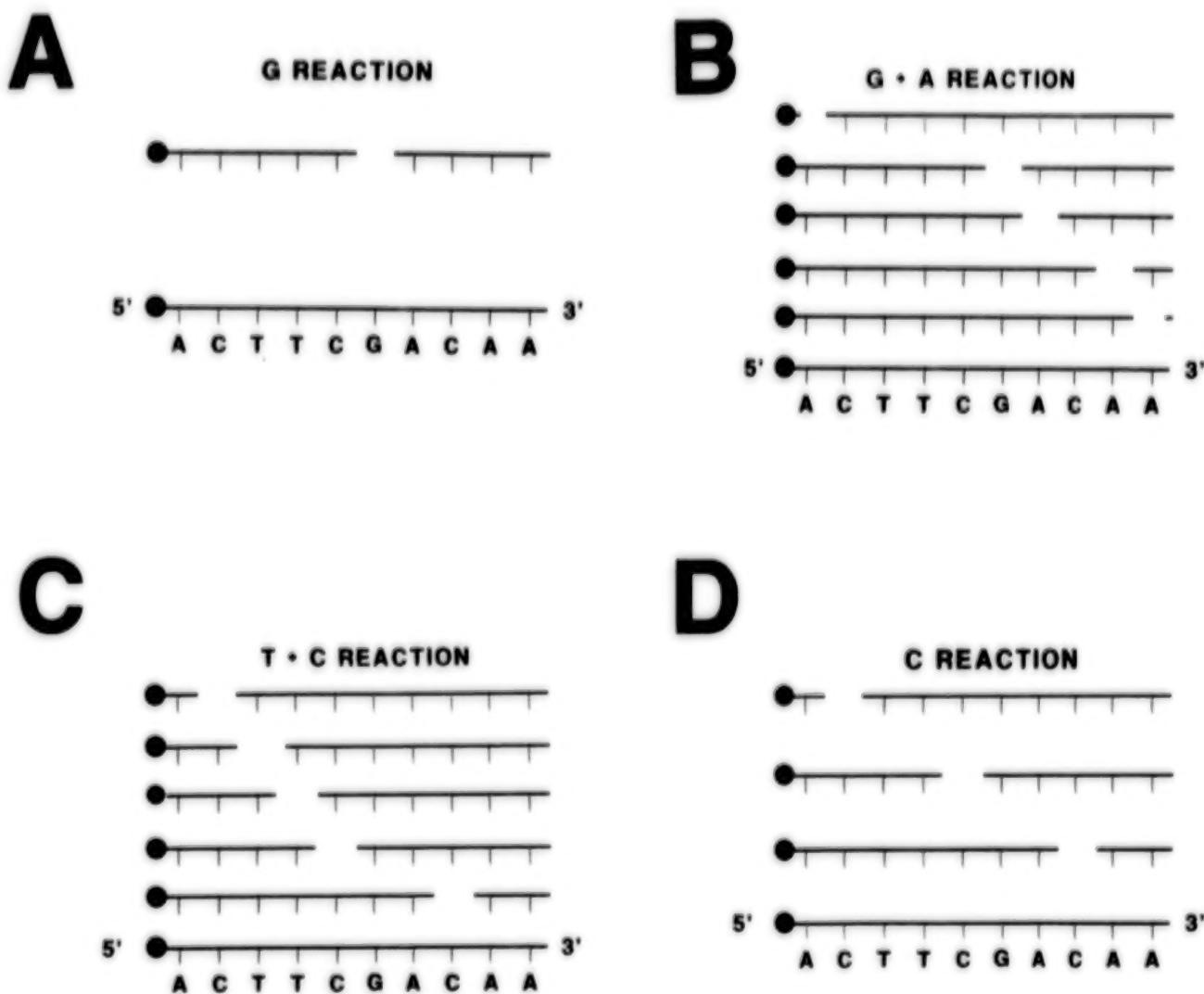
Figure 2. Maxam and Gilbert sequencing reaction products. The DNA fragments resulting from the four chemical degradation reactions used in Maxam and Gilbert sequencing are shown. In each panel, the original DNA sequence is shown at the bottom, and the family of fragments obtained from limited cleavage of the sequence are shown above. The circle represents the radioactive label. Panel A: Products resulting from the G reaction. Panel B: Products resulting from the G+C reaction. Panel C: Products resulting from the T+C reaction. Panel D: Products resulting from the C reaction.

Following isolation of a single or double-stranded DNA molecule with a label only on one end, the DNA is aliquoted into four separate tubes. In one tube (the G reaction), guanines (G) are chemically modified by dimethylsulphate. In another (G+A reaction), guanines and adenines are modified by acid. In the third (T+C reaction), thymines and cytosines are modified by hydrazine. In the fourth (C reaction), cytosines are modified by hydrazine and salt. In each reaction, the purine or pyrimidine rings of the bases are modified or removed (Other chemical treatments also have been described for modification of bases by Hindley 1983). Subsequent treatment of the DNA with piperidine causes removal of the modified base, breaking the DNA strand at the place where the modified base resides. The reaction conditions are adjusted so that on average each DNA strand is cleaved only one time. Thus, for example, the G reaction yields a family of DNA fragments, with the ends of each corresponding to places where guanines were present in the original sequence. Figure 2 shows the fragments generated from the degradation of a DNA sequence in the four reactions.

To fractionate the fragments by size, the four reactions are loaded onto adjacent lanes of a DNA sequencing gel (Sanger and Coulson 1978). The DNA sequencing gels have the capability to resolve DNA molecules differing in length by a single base. They are typically made of polyacrylamide and contain urea, which denatures DNA, allowing the single stranded labelled products to be resolved without the interfering effects of secondary structure formation. Thus, the single stranded DNA molecules migrate in proportion to their length.

Following electrophoresis and autoradiography of the gel, the sequence can be read by identifying which reaction produced each successively larger band. Figure 3 shows the pattern of bands obtained for a DNA sequence. When reading a DNA sequence, the orientation, or 5' to 3' direction of the sequence, is determined by the position of the radioactive label on the DNA sequence. If the 5' end was labelled, the sequence will read from the 5' to 3' direction going from the bottom to the top of the gel, because the smaller fragments, which are at the bottom of the gel, are closest to the 5' label. Conversely, if the 3' end was labelled, the sequence will be read 3' to 5', because the smallest fragments are closest to the 3' label.
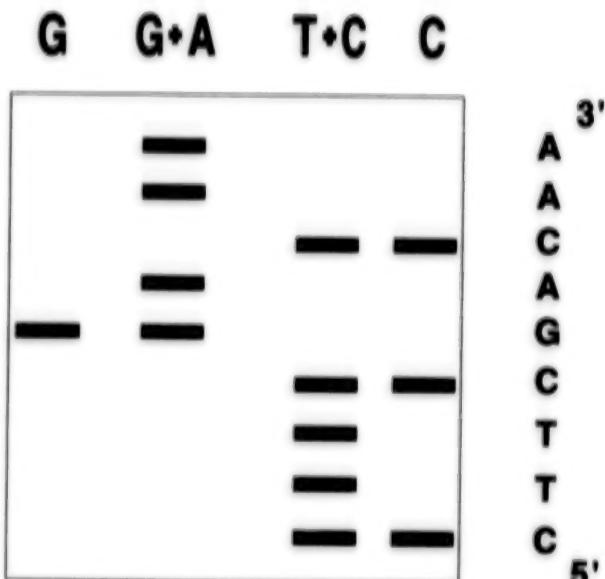
Figure 3. Banding pattern obtained from a Maxam and Gilbert sequencing gel. The sequence obtained from the Maxam and Gilbert sequencing of the DNA sequence shown in Figure 1 is shown. The sequence is listed to the right of the gel, with the 5' to 3' direction indicated.

Sanger Dideoxynucleotide Sequencing

The dideoxynucleotide chain termination method of DNA sequencing (Sanger et al. 1977) is based on elongation by enzymatic synthesis of a short DNA strand hybridized to a template strand.

As with any reaction in which DNA synthesis is carried out, the Sanger method of sequencing requires a single-stranded DNA template (which is the DNA to be sequenced), an oligonucleotide (generally 15-17 nucleotides in length), which is homologous to a region of the DNA template and referred to as a primer, the deoxynucleotide components of DNA (dATP, dCTP, dGTP, and dTTP), and a DNA polymerase. Because the primer must hybridize to one particular site flanking the region of interest, some sequence information must be known about the template. This can be accomplished in one of two ways. The template can be cloned into a sequencing vector which has a universal primer site flanking the inserted DNA. Primers which are complementary to this site can be purchased or synthesized. Alternatively, if part of a DNA sequence is already known, a primer can be synthesized that is complementary to that sequence and can be used to prime DNA synthesis of the unknown DNA sequence.

In the first step of the sequencing reaction, the primer is hybridized to the template. The hybridization reaction products are then aliquoted into four separate tubes. Each tube has, in addition to
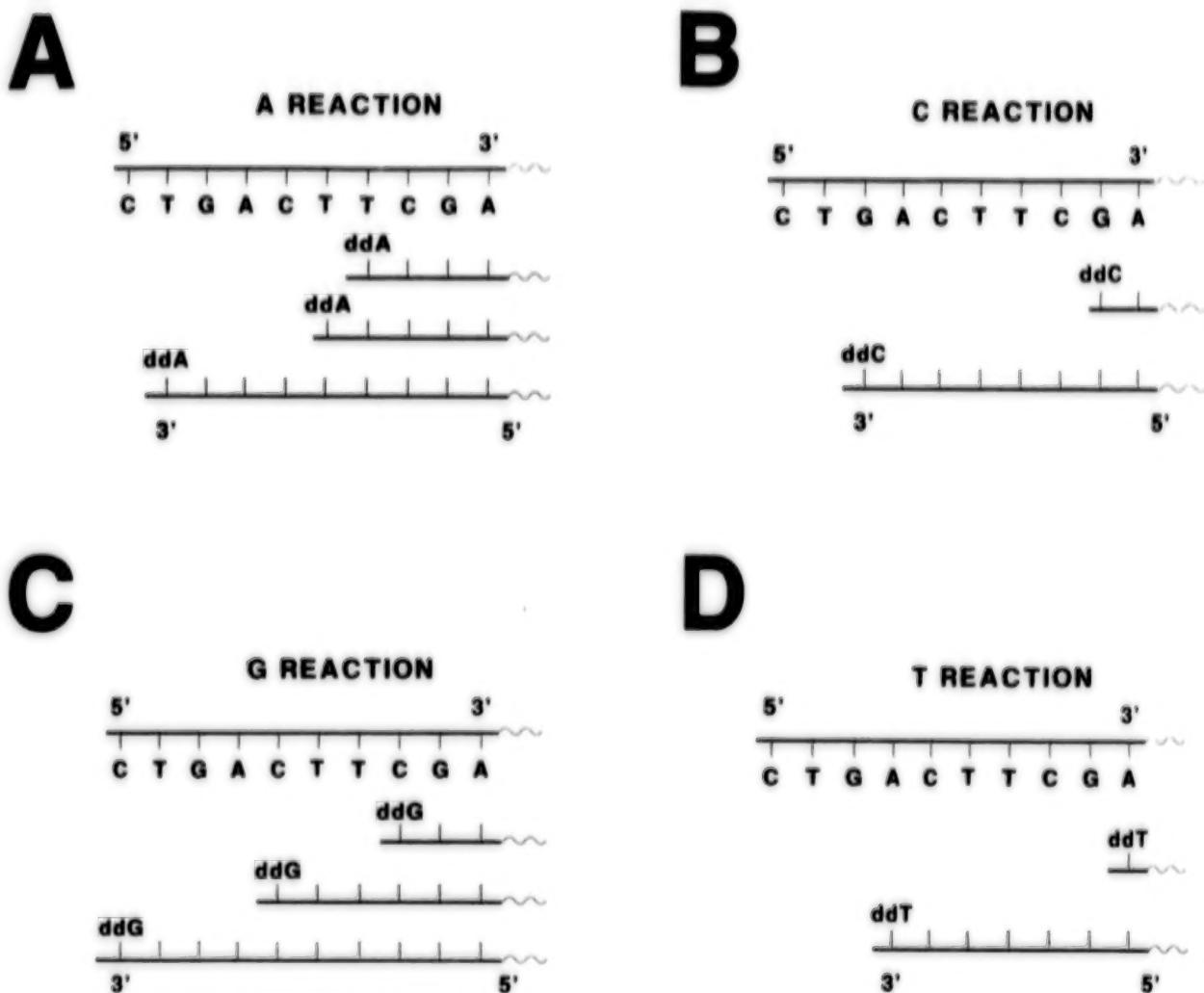
Figure 4. Sanger sequencing reaction products. The DNA fragments resulting from the four sequencing reactions used in Sanger sequencing are shown. In each panel, the original DNA sequence is shown at the bottom, and the family of terminated extension products are shown above. Panel A: Products from the A reaction. Panel B: Products from the C reaction. Panel C: Products from the G reaction. Panel D: Products from the T reaction.

the four deoxynucleotides (one of which is radioactively labeled) and a DNA polymerase, a chain terminator. The chain terminators most commonly used are 2′,3′-dideoxynucleotides (ddNTPs). Each reaction tube contains only one type of chain terminator, either ddATP, ddCTP, ddGTP, or ddTTP. These reactions are referred to as the A reaction, C reaction, G reaction, and T reaction, respectively. The dideoxynucleotides are recognized by the polymerase and incorporated into the growing DNA strand. However, because the 3′ position of each ddNTP has a hydrogen rather than a hydroxyl group, there is no site for covalent binding by an additional nucleotide. Since a limited amount of chain terminators is used (they are present at approximately one-one hundredth the concentration of the deoxynucleotides), the ddNTPs are only occasionally incorporated into

the growing DNA strand. Thus, for example, in the A reaction, ddATP molecules are occasionally incorporated rather than dATP molecules, terminating further synthesis of that strand. Following DNA synthesis, each tube contains a family of DNA molecules, each terminating with the particular ddNTP in that tube (Figure 4). In the A reaction, each DNA molecule terminates with a ddATP residue. Because the reaction products are complementary to the template, each ddATP corresponds to each place there was a T in the template DNA. Similar reaction products are obtained for the C, G and T reactions, each terminating with the appropriate ddNTP. The DNA fragments obtained from the four reactions are loaded into adjacent lanes on a sequencing gel. After electrophoresis and autoradiography of the gel, the sequence can be read from the autoradiogram by

identifying the position of each successively larger band. An example of the pattern of bands obtained from a DNA sequence is shown in Figure 5. The sequence is read in a 5' to 3' direction, because DNA synthesis always occurs 5' to 3'. Since the smallest fragments are closest to the 5' end of the synthesized DNA, the sequence is read from the bottom to the top of the autoradiogram.

The Sanger method of sequencing was described using the large fragment of polymerase I from *E. coli*, also referred to as the Klenow fragment (Sanger *et al.* 1977). More recently, other polymerases have been used for Sanger sequencing, which avoid some problems which can arise using Klenow (Tabor and Richardson 1987). A modified polymerase from the bacterial virus T7, called Sequenase (United States Biochemical Corporation, Cleveland, OH), can synthesize through certain sequences which Klenow cannot synthesize well (for example, stretches of cytosines sequence better with Sequenase) (Tabor and Richardson 1987). *Taq* polymerase, from the thermalphilic bacterium *Thermus aquaticus* (Chien *et al.* 1976), which is the enzyme used in the Polymerase Chain Reaction (PCR), can be used for sequencing (Peterson 1988; Innis *et al.* 1988). Because the synthesis reaction using *Taq* polymerase can be performed at a higher temperature (72 °C) than is used with Klenow or Sequenase (37 °C), the possibility of the template folding back on itself and undergoing some base
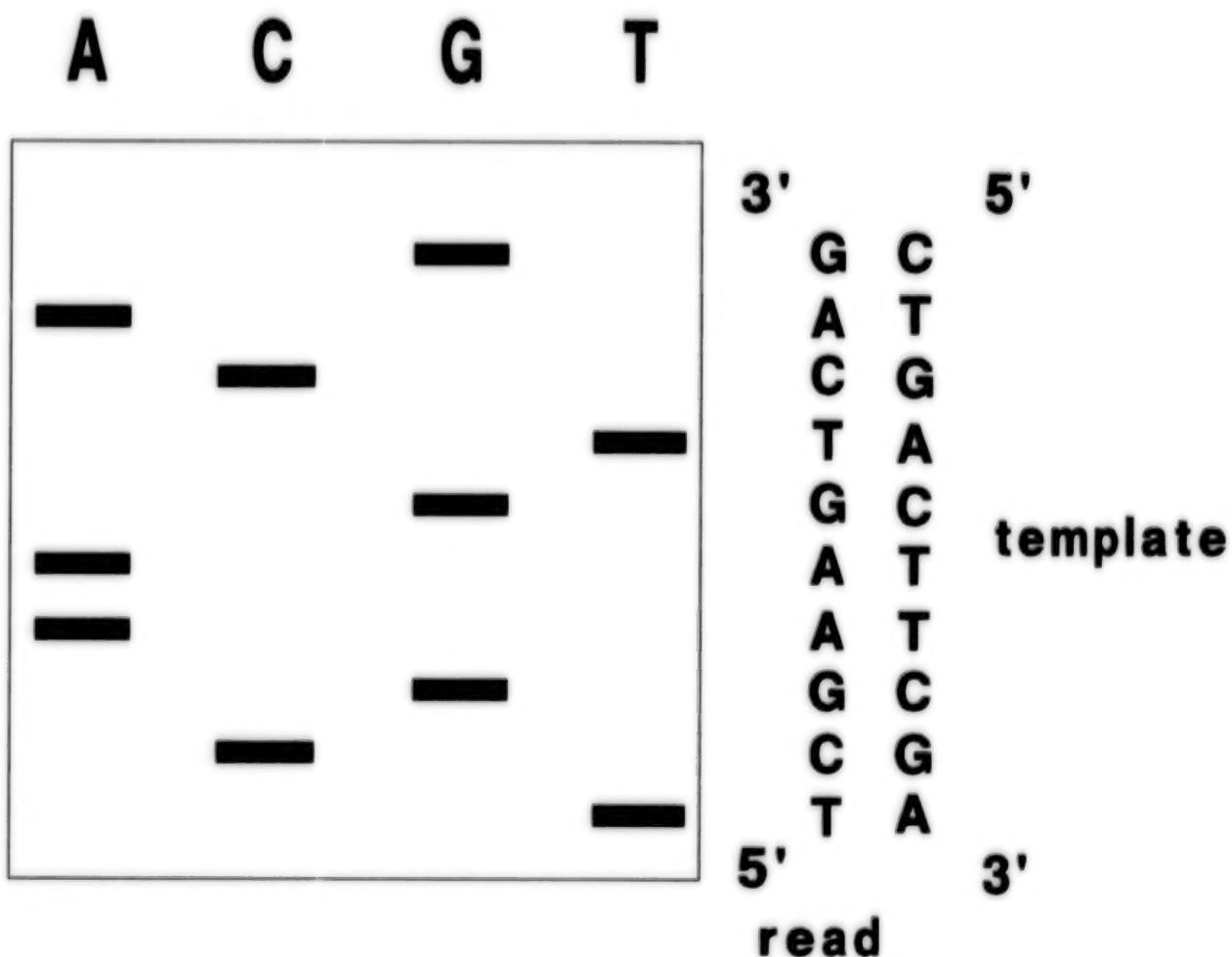
Figure 5. Banding pattern obtained from a Sanger sequencing gel. The sequence obtained from the Sanger sequencing of the DNA sequence shown in Figure 4 is shown. The sequence as it is read from the gel is shown directly to the right of the gel, while the sequence of the template sequence (the complementary sequence) is shown at the far right. The 5' to 3' direction of the sequence as it is read from the gel, as well the orientation of the template sequence, is indicated.

pairing (that is, forming secondary structures), which interferes with DNA synthesis, can be avoided. Therefore, some sequences which are difficult to sequence using Klenow and Sequenase can be more readily sequenced using *Taq*. Additionally, direct sequencing of PCR-amplified DNA can be done using *Taq*, eliminating the need for multiple polymerases.

COMPARISON OF MAXAM AND GILBERT AND SANGER SEQUENCING METHODS

Any purified DNA fragment can be sequenced by the Maxam and Gilbert method. However, for Sanger sequencing, because primers are necessary, the DNA of interest must either be cloned into a sequencing vector, which has a universal primer site (Waye 1990), or alternatively, a primer known to be complementary to part of the sequence adjacent to the sequence of interest must be synthesized. While either double-stranded or single-stranded DNA can be sequenced using the Maxam and Gilbert method (providing the DNA has only one radioactive label on it), sequencing of double-stranded DNA using the Sanger method is difficult (Deininger 1983; Guo *et al.* 1983). This is because the complementary DNA strand competes with the primers for binding to the template during hybridization. However, the Sanger method is a simpler reaction to carry out, and it does not use the labile and toxic chemicals required for Maxam and Gilbert sequencing.

Because Maxam and Gilbert uses end-labelled DNA (that is, one label per DNA molecule), P^{32} must be used to obtain a strong enough signal. Using the Sanger method, radiolabelled nucleotides are incorporated throughout the reaction products. Thus, S^{35}-dATP can be used as the isotope even though it is of lower energy. With the use of S^{35}, reaction products are stable for several days when frozen, and the use of S^{35} yields sharper (narrower) bands on the autoradiograms (Biggin *et al.* 1983).

AUTORADIOGRAPHY OF SEQUENCING GELS

Radioactive bands are detected by autoradiography. After the products of sequencing reactions are separated by gel electrophoresis, the gel must be prepared for exposure to X-ray film. Although X-ray film may be applied directly to a wet gel covered in plastic wrap (Hindley 1983), fixing and drying of the gel prior to autoradiography improves signal intensity and band sharpness (Biggin *et al.* 1983). Urea quenches the signal of the iso-

tope, can precipitate out, and makes the gel more difficult to handle. Therefore, gels are washed in a solution of acetic acid and methanol (Biggin *et al.* 1983). This fixes the DNA fragments in the gel and serves to remove urea from the gel. Gels are then placed on absorbent paper or an old piece of X-ray film and dried by placement on a gel dryer or in an oven. The X-ray film can then be laid directly on the dried gel. If S^{35} is used, the gel must be fixed and dried. Plastic wrap is not used because it blocks the decay product of S^{35}, greatly increasing the required exposure time.

ANALYSIS OF SEQUENCING GELS

Once autoradiograms of the sequencing gels are obtained, the DNA sequence must be read from the films. The sequence can be determined and recorded manually. However, automated gel readers are available. The use of these reduces the possibility of errors that can occur during the transcription of data. One type of gel reader (for example, the IB Gel Reader, sold by International Biotechnologies, Inc., New Haven, CT) consists of a light box with a stylus attached, which is connected to a computer. The autoradiogram is placed on the light box, and the stylus allows the user to define the lanes and record the band pattern. The sequence is then recorded directly into the computer. Other types of gel readers (for example, the Gel-Read[R] Semi-Automatic DNA Sequence Reader, sold by Bio-Rad Laboratories (Richmond, CA); see Elder *et al.* 1986) optically scans the autoradiogram, and a software program finds lanes and assigns bases.

Since sequence data collected and stored manually quickly becomes unmanageable, DNA sequence analysis software is desirable for the storage and analysis of sequence data. Available software programs (such as MicroGenie, from Beckman Instruments, Inc., (Palo Alto, CA); see Queen and Korn 1984; and IBI/Pustell, from International Biotechnologies, Inc.), in addition to storing data, can find homologies in other DNA sequences. These homologies can be in DNA sequences previously sequenced in the laboratory or to sequences in a DNA data bank such as GenBank[R] (Bilofsky and Burks 1988). The programs can find restriction enzyme recognition sites, thus creating restriction maps, and DNA signal sequences such as transcription initiation sites and intron/exon junction sites. Additionally, they can find possible coding regions of DNA (open reading frames) and predict putative protein sequences.

The DNA sequences must be confirmed. It is not sufficient to sequence one strand of a DNA

fragment one time. Complementary strands of a DNA fragment must be sequenced several times. Additionally, overlapping DNA fragments, in other words, DNA fragments which contain some of the same sequence in addition to sequences adjacent to the particular sequence, must be sequenced. Sequences at the top of a gel are compressed and difficult to read; thus, by sequencing overlapping fragments, a sequence that was at the top of the gel from one fragment can be moved to a superior resolving region of the gel if an overlapping fragment is sequenced. This sequence confirmation reduces errors in reported DNA sequences (although errors in reported DNA sequences occur due to multiple factors, including misincorporation of nucleotides by polymerases and misinterpretation of data; see Krawetz 1989 for a discussion of sequence errors).

AUTOMATION OF DNA SEQUENCING

Automated DNA sequencers allow rapid, automatic detection of sequencing reaction products as they are separated, eliminating errors due to manual transcription of sequencing data. There are two systems based on the use of fluorescent, rather than radioactive labels, which are being used by a number of researchers and will be discussed here. One is manufactured by Automated Biosystems, Inc. (ABI) and the other, the Genesis II, is made by E. I. DuPont de Nemours and Co.

ABI System

The ABI automated DNA sequencing system is designed for use with the Sanger method of sequencing. The approach is similar to that described above, except that the primer is labelled with one of four different fluorescent dye molecules (Smith *et al.* 1985, 1986). No radioactive nucleotides are used. The primer annealing step is done in four different tubes. After the extension of the primers by the polymerase, each molecule will have a ddNTP at its 3′ end and a fluorescent dye at the 5′ end. In each reaction, the primer has a different label, designating the appropriate ddNTP. Because each set of reaction products has a different fluorescent tag, the reaction products can all be loaded in a single well of a gel in the ABI system. During electrophoresis, the DNA molecules pass a fixed point at which a laser beam excites the dye molecules. Photomultiplier tubes and filters detect the appropriate emitted signal. The signals are recorded by the computer and translated to the corresponding bases. Since each dye emits light at a different wavelength, and each dye corresponds to one particular ddNTP, a DNA sequence can be generated.

DuPont System

The Genesis II DNA Sequencer by DuPont also uses fluorescent dyes. However, rather than being attached to the primers, the dyes are attached to the ddNTPs (Prober *et al.* 1987). One dye is attached to ddATP, a different one to ddCTP, and similarly for ddGTP and ddTTP. Thus, the Sanger annealing and extension reactions for sequencing can be carried out in a single tube, and, in contrast to the ABI system described above in which particular labelled primers are required, any primer can be used. DNA molecules ending in each of the different ddNTPs will have different dyes attached. Each reaction is run in a single lane in the DuPont apparatus, and (as in the ABI system), as the DNA molecules pass a fixed point, the dyes are stimulated by a laser and the emitted wavelengths pass through filters and are recorded by photomultiplier tubes. The pattern of emitted wavelengths is recorded by the computer and translated to a DNA sequence.

Advantages of Automated DNA Sequencing

The major advantage of automated sequencing is the speed at which data can be collected. There is no lag time between electrophoresis and data collection, which occur simultaneously. Errors arising from manual transcription of data are avoided. The use of automated DNA sequencers also avoids the post-electrophoresis manipulation of gels required for manual sequencing. Additionally, the use of radioactive isotopes is avoided. Since the reactions are run in one lane per DNA sequence, rather than four, more samples per gel can be analyzed.

FORENSIC APPLICATIONS OF DNA SEQUENCING

Theoretically, any region of DNA that demonstrates polymorphism can be exploited for forensic purposes. By comparing the sequence of a polymorphic region from DNA from different forensic samples, one can genetically characterize them.

The D-loop region of mitochondrial DNA, quite variable among individuals, is a DNA sequence that may prove useful for forensic analysis (see Rath and Merril 1990, for a thorough discussion of this). Briefly, mitochondria are intracellular organelles with their own genome, separate from the nuclear genome. There are many copies of the mitochondrial genome per cell, so any region of

mitochondrial DNA is present in many more copies per cell than is any particular region of nuclear DNA, which is present in only two copies per cell. Thus, detection of mitochondrial sequences is easier than detection of nuclear DNA. Since DNA quantity is often limited in forensic samples, this is an important consideration.

However, sensitivity of detection can be enhanced by the amplification of DNA by PCR (Saiki *et al.* 1985, 1988; Erlich *et al.* 1990). The amplified DNA can then be directly sequenced (Peterson 1988; Innis *et al.* 1988). If the DNA is cloned into sequencing vectors, identification and isolation of specific DNA can take several days to weeks, whereas this can be done in one to two days using PCR. Thus, in two to three days, DNA from a small amount of starting material can be amplified and sequenced.

However, several problems must be overcome if DNA sequencing is to become a useful forensic tool. First, regions sufficiently polymorphic must be identified, and population data on frequencies of polymorphisms must be obtained. Proper statistical methods for analyzing sequence polymorphisms must be employed. Second, it must be ensured that errors in sequencing are minimized. This will require, at a minimum, sequencing of both strands of a DNA fragment to confirm sequence data. A third problem which must be addressed is the problem of sequencing DNA from mixed body fluids. It could be difficult to interpret and assign signals arising from multiple bases at one position. Finally, DNA sequencing as a forensic procedure must be properly validated (for example, studies of the effects of contamination by chemicals, microorganisms, and other human biological specimens on sequencing).

CONCLUSION

Sequencing of DNA is a powerful tool which detects polymorphisms at their most fundamental level—the order of the nucleotides in DNA. While the technique is cumbersome and requires a good deal of expertise to perform it, advances in technology which simplify the procedure and improve the automation make it possible that DNA sequencing will be useful to forensic scientists. The future of DNA sequencing in forensic science depends on continued technological improvements, and on the discovery of informative polymorphic regions, and, on the appropriate statistical analysis of these polymorphisms.

REFERENCES

Biggin, M. D., Gibson, T. J. and Hong, G. F. (1983). Buffer gradient gels and S[35] label as an aid to rapid DNA sequence determination, Proc. Natl. Acad. Sci. USA 80:3963–3965.

Bilofsky, H. S. and Burks, C. (1988). The GenBank[R] genetic sequence data bank, Nucleic Acids Res. 16:1861–1863.

Chien, A., Edgar, D. B. and Trela, J. M. (1976). Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*, J. Bacteriol. 127:1550–1557.

Deininger, P. L. (1983). Approaches to rapid DNA sequence analysis, Anal. Biochem. 135:247–263.

Elder, J. K., Green, D. K. and Southern, E. M. (1986). Automatic reading of DNA sequencing gel autoradiographs using a large format digital scanner, Nucleic Acids Res. 14:417–424.

Erlich, H. A., Higuchi, R., von Beroldingen, C. and Blake, E. (1990). The Use of the Polymerase Chain Reaction for Genetic Typing in Forensic Samples, In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C.

Guo, L-H., Yang, R. C. A. and Wu, R. (1983). An improved strategy for rapid direct sequencing of both strands of long DNA molecules cloned in a plasmid, Nucleic Acids Res. 11:5521–5540.

Hindley, J. (1983). DNA Sequencing. Elsevier Biomedical Press, Amsterdam.

Hyman, E. D. (1988). A new method of sequencing DNA, Anal. Biochem. 174:423–9436.

Innis, M. A., Myambo, K. B., Gelfand, D. H. and Brow, M. A. D. (1988). DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA, Proc. Natl. Acad. Sci. USA 85:9436–9440.

Krawetz, S. A. (1989). Sequence errors described in GenBank: a means to determine the accuracy of DNA sequence interpretation, Nucleic Acids Res. 17:3951–3958.

Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing DNA, Proc. Natl. Acad. Sci. USA 74:560–564.

Maxam, A. M. and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavage, Methods Enzymol. 65:499–560.

Peterson, M. G. (1988). DNA sequencing using Taq polymerase, Nucleic Acids Res. 16:10915.

Prober, J. M., Trainor, G. L., Dam, R. J., Hobbs, F. W., Robertson, C. W., Zagursky, R. J., Cocuzza, A. J., Jensen, M. A. and Baumeister, K. (1987). A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides, Science 238:336–341.

Queen, C. and Korn, L. J. (1984). A comprehensive sequence analysis program for the IBM personal computer, Nucleic Acids Res. 12:581–599.

Rath, D. S. and Merril, C. R. (1990). Mitochondrial DNA and its Forensic Potential. In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn G. T., Erlich, H. A. and Arnhein, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia, Science 230:1350–1354.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, Science 239:487–491.

Sanger, F. and Coulson, A. R. (1978). The use of thin acrylamide gels for DNA sequencing, FEBS Letters 87:107–110.

Sanger, F., Niklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. USA 74:5463–5467.

Smith, L. M., Fung, S., Hunkapiller, M., Hunkapiller, T. and Hood. L. E. (1985). The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis, Nucleic Acids Res. 13:2399–2412.

Smith, L. M., Sanders, J., Kaiser, T., Hughes, P., Dodd, C. and Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis, Nature 321:674–679.

Tabor, S. and Richardson, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase, Proc. Natl. Acad. Sci. USA 84: 4767–4771.

Waye, J. S. (1990). Cloning and Recombinant DNA Technologies for the Development of Hybridization Probes. In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C.

BLANK PAGE

MITOCHONDRIAL DNA AND ITS FORENSIC POTENTIAL

Denise S. Rath and Carl R. Merril

National Institute of Mental Health
Washington, D.C. 20032

The purpose of this paper will be to examine the advantages of using the mitochondrial genome for forensic analysis and to present some of the advanced technologies which are applicable for this work. Recent data on the mitochondrial DNA and its association with disease states will also be presented.

Recent years have seen an explosion of technology aimed at deciphering the information hidden in the human genome. This data, once revealed, could lead to a detailed understanding of the molecular biology of man. Additionally, the ability to read genomic information would permit the differentiation of individuals at the DNA level and allow the interpretation of differing phenotypes. This capacity would also be invaluable to forensic scientists in their attempts to individualize forensic samples. The complete sequencing of the nuclear genome is probably decades away, however, the mitochondrial DNA, due to its vastly smaller size (16,596 base pairs (bp)), has been sequenced in its entirety and all of its genes have been mapped.

In addition to the small size of the mitochondrial genome and its relatively simple gene organization, the DNA has been shown to have a higher rate of nucleotide sequence evolution than the nuclear DNA (Brown *et al.* 1979, 1982; Cann *et al.* 1987). The higher rate of evolution can be attributed to several factors: 1) the apparent lack of efficient mtDNA repair enzymes (Miyaki *et al.* 1977; Fukunaga and Yielding 1979), 2) the poor fidelity of the mtDNA polymerase (Kunkel and Loeb 1981) and 3) an exposure to endogenous mutagenic agents, such as oxygen free radicals and lipid peroxides, which are a by-product of oxidative phosphorylation.

GENETICS

A human being is made up of approximately 3.5×10^{13} cells. Since each of these cells contains between 100 and 1,000 mitochondria, an average human has 1.7×10^{16} mitochondria (Robin and Wong 1988). As mitochondria contain multiple copies (2–3) of a small, circular, double-stranded DNA molecule, then each human has approximately 1×10^{17} copies of their mitochondrial genome.

This is a substantial number, one hundred thousand trillion. It would be hard for any creature carrying so many copies of a genome not to leave a few copies whenever they move about.

The human mitochondrial genome consists of 16,596 bp and encodes a set of transfer RNAs, two ribosomal RNAs and 13 proteins. All of these proteins play a role in the mitochondrial electron transport chain.

Two of the many characteristics of mitochondrial DNA which distinguish this genome from the nuclear genome are its monoclonal nature in each individual and its matrilineal route of inheritance. Hutchison *et al.* (1974) were the first to demonstrate maternal inheritance of the mtDNA in mammalian species by using restriction endonuclease analysis on domesticated equines. They found that the enzyme digest patterns for horses differed from those of a donkey. They used this information to examine the mtDNA for existing crosses between the two species, mules and hinnies, and showed that the mtDNA pattern of the offspring always corresponded to the maternal pattern.

Giles *et al.* (1980) used the same technique to expand this analysis to human mitochondrial DNA. They examined the restriction patterns from mtDNA in peripheral blood platelets in seven pedigrees. In each family that exhibited differences in the maternal and paternal mtDNA, the children were all shown to display the maternal type. Additionally, Case and Wallace (1981) using restriction analysis on fibroblasts from a single pedigree illustrated that there was no paternal contribution and verified that mitochondrial DNA was maternally transmitted.

Due to its unique mode of inheritance, the mitochondrial DNA exhibits an apparent lack of recombination. This fact combined with the poor fidelity of the mtDNA polymerase and absence of mtDNA repair mechanisms lead to a higher rate of mutation fixation in the mitochondrial genome compared with nuclear DNA. Certain regions of the mitochondrial genome appear to be evolving at 5 to 10 times the rate of single copy nuclear genes (Brown *et al.* 1979, 1982; Cann *et al.* 1987). The higher rate of sequence evolution has proved useful in genealogy and evolutionary studies and it should

prove to be valuable in forensic studies. Such a high rate of evolution may provide the ability to differentiate individuals even within multi-generational families (Hausworth *et al.* 1984).

The potential to use mitochondrial DNA to distinguish human samples was first demonstrated by Wesley Brown (1980) in a study of restriction fragment length polymorphisms (RFLPs). By using 18 restriction enzymes, he was able to characterize samples from 21 individuals by their enzyme digest patterns. To date, many studies (Brown and Goodman 1979; Denaro *et al.* 1981; Ferris *et al.* 1981; Blanc *et al.* 1983; Aquadro and Greenburg 1983; Cann *et al.* 1984, 1987; Horai *et al.* 1984; Horai and Matsunuga 1986; Bonne-Tamir *et al.* 1986; Brega *et al.* 1986 a,b; Harihara *et al.* 1986; Scozzari *et al.* 1988; Minshu *et al.* 1988; Santachiara-Benerecetti *et al.* 1988) have examined human mtDNA polymorphisms and have uncovered patterns which correspond to ethnic and racial differences as well as a tremendous variation within every given sample population. This information has been used to establish evolutionary trees for the Homo sapiens species and could become invaluable to forensic science.

However, one of the concerns in applying this data to the field of forensics is the homogeneity of the mitochondrial DNA within an individual. Potter *et al.* (1975) used mtDNA from several mammalian species to examine the consistency of restriction patterns in a variety of tissues from individual animals. They found no differences between organs and concluded that mammalian mitochondrial DNA is monoclonal in origin.

Even more convincing evidence for a monoclonal nature of the mitochondrial DNA in each individual was provided by the studies of Monnat and several colleagues (Monnat *et al.* 1985; Monnat and Loeb 1985; Monnat and Reay 1986) in which mitochondrial DNA fragments derived from various tissues from several individuals were cloned into bacteriophage M13. Individual clones were subsequently sequenced. These investigators sequenced up to 81 clones (17,527 nucleotides) from a single patient and found only two within-individual differences (nucleotide substitutions). They concluded that there is a high degree of mtDNA sequence homogeneity in somatic cells which suggests the existence of a mechanism that limits the development of nucleotide sequence divergence in human mitochondrial DNA.

More recent studies, however, have discovered that mitochondrial DNA heteroplasmy does exist in man. Holt *et al.* (1988b) performed restriction enzyme analysis on mitochondrial DNA isolated from leukocytes and muscle from twenty-five patients with mitochondrial myopathies and their relatives. Digestion of the leukocyte mtDNA with multiple enzymes revealed no differences between affected and unaffected individuals within a single maternal line. However, when muscle mtDNA was analyzed, nine of the twenty-five patients were found to have two populations of mtDNA, one of which contained deletions of up to 7 Kilobases (Kb) in length. This was the first conclusive evidence of mtDNA heteroplasmy in humans as well as the first association of a mtDNA defect with disease in man. These findings were corroborated by Johns *et al.* (1989), who found a dimorphic population of deleted mtDNAs in skeletal muscle, and Poulton *et al.* (1989), who discovered abnormal mtDNAs with (about 8 Kb) duplications present in several tissues, in related clinical diseases. Poulton and her co-workers expanded the findings to include tissue mosaicism. They felt that different ratios of the two mitochondrial genotypes in different tissues could explain the varying phenotypes of individual patients.

Despite these recent clinical findings, it appears that heteroplasmy of the mitochondrial DNA is a relatively rare event and associated only with a limited number of maternally inherited diseases. Therefore, these few cases should not preclude the usage of mtDNA for forensic analysis.

METHODOLOGIES

Restriction Fragment Length Polymorphisms (RFLPs)

The bulk of the information obtained from the mitochondrial DNA to date had been derived from restriction endonuclease analysis and Southern blotting. Historically, this was one of the first molecular biological methodologies available. This technique employs restriction enzymes isolated from bacterial strains which recognize specific nucleotide sequences, usually 4–6 nucleotides, in the DNA. These enzymes cut the DNA at these sites and the cleavage patterns are then analyzed on agarose gels. The resulting bands can be visualized directly on the gel with ethidium bromide or the DNA fragments can be transferred to nitrocellulose paper (Southern blotting) (Southern 1975), hybridized with a radioactive probe, and observed following autoradiography.

This technique has already provided a substantial amount of information about the mitochondrial genome, such as RFLPs which correspond to spe-

cific ethnic or racial groups, which is of relevance to the forensic sciences. However, this method is rather limited in that it can only examine a very small percentage of the genome. Using 7 restriction enzymes with tetranucleotide recognition sites, Brown (1980) estimated that only 5.3% of the mitochondrial genome was being screened. Although this methodology can detect most insertions and deletions, restriction endonuclease mapping is particularly insensitive when nucleotide sequence differences are distributed randomly. For example, restriction enzyme analysis could fail to detect randomly distributed nucleotide sequence changes that involve as many as 10% of all nucleotide positions in a population of DNA molecules (Monnat and Loeb 1985). Additionally, this method of DNA analysis requires microgram amounts of relatively undegraded DNA for multiple enzyme digestions. Such DNA frequently cannot be obtained from forensic samples.

RNA–DNA and DNA–DNA Duplexes

Many base substitutions cannot be evaluated by restriction endonuclease analysis since they do not alter a restriction site. As a result of these limitations, alternate methodologies have been developed to detect single bp mutations in cloned or genomic DNA. Myers et al. (1985) published a method which involves the enzymatic cleavage of RNA at a single base mismatch in an RNA–DNA hybrid. This technique utilizes RNA probes which are synthesized by SP6 RNA Polymerase (Melton et al. 1984) and then hybridized to cloned or genomic DNA. Ribonuclease A is then used to cleave mismatches in the RNA–DNA hybrids. This method can detect 70% of these mismatches if sense and anti-sense probes are used. Obviously, this technique would also detect small insertions or deletions which could be missed in restriction enzyme mapping.

As this method does not detect all mutations, Cotton et al. (1988) evaluated the chemical reactivity of mismatched bases in DNA–DNA heteroduplexes. Radioactively labeled DNA probes were prepared from M13 clones and hybridized with unlabeled DNA to form heteroduplexes. The heteroduplex DNAs were incubated with osmium tetroxide, for cytosine mismatches, followed by treatment with piperidine to cleave the DNA at the modified mismatched base. By using sense and anti-sense probes, all potential mismatches can be examined by this method.

The value of these particular technologies is that they allow evaluation of a relatively large segment of the genome (about 1 Kb) in any given experiment. This ability is certainly applicable to the mitochondrial DNA since only about a dozen probes would be necessary to screen the entire genome for base substitutions, deletions and insertions. If a given region, such as the D-loop, proves to be hypervariable enough to determine individuality, then only a single probe may provide sufficient resolution for forensic analysis.

DNA Sequencing

Two DNA sequencing methodologies were published in 1977: the chemical cleavage method developed by Maxam and Gilbert (1977) and the chain termination technique invented by Sanger et al. (1977). The Sanger method is the more generally used procedure. The procedure uses an oligonucleotide primer which is annealed to the purified DNA fragment to be sequenced. This primer is extended along the template by a DNA polymerase using the four deoxyribonucleotides (dATP, dCTP, dTTP and dGTP) and one of the four dideoxyribonucleotides (dATP, dCTP, dTTP or dGTP). When a dideoxyribonucleotide is incorporated into the sequence, chain elongation is terminated. In this manner, fragments of different lengths are generated depending upon the site of dideoxyribonucleotide insertion. This reaction is performed on four test tubes, each tube containing a different deoxyribonucleotide.

Since one of the deoxyribonucleotides is usually radiolabeled with S^{35} or P^{32}, the fragments can be fractionated on a denaturing polyacrylamide gel and visualized by autoradiography. The result is a ladder of DNA bands differing by only one nucleotide which can be read from lane to lane to determine the DNA sequence.

While this method can be extremely informative, it is also highly labor intensive. Each reaction can usually only give the sequence for 100–250 bp. This is then multiplied by the length of the region of interest and the number of samples to be analyzed. The technique also requires either knowledge of a partial sequence in order to prepare the oligonucleotide primers or a cloning step into a vector with commercially available primers. However, it should be noted that a considerable effort is being expended to automate these procedures and a number of companies are now marketing automated DNA sequencing machines.

Polymerase Chain Reaction

The polymerase chain reaction (PCR) (Saiki *et al.* 1985, 1986, 1988; Mullis *et al.* 1986; Mullis and Faloona 1987; Li *et al.* 1988) uses Taq polymerase to synthesize multiple copies of a DNA segment whose ends are specified by oligonucleotide primers which have been added to the reaction mixture. The DNA segment of interest may only be a tiny fraction of the starting material and one can even study DNA in a crude cell lysate (Saiki *et al.* 1986) with this technique. DNA is synthesized by multiple repetitions of a three stage cycle which consist's of denaturation, annealing and extension. Initially, the two strands of the DNA segment are separated by heating to an elevated temperature (94 °C). On cooling, renaturation of the DNA strands is impaired by an excess of the two primers which anneal to the parent strands. The primers are short, single stranded oligomers of known DNA sequence which are homologous to regions that flank the DNA sequence to be amplified. DNA is then synthesized along the parent strands by a thermostable polymerase (Taq polymerase) primed by the annealed oligonucleotide segments. By repeating the cycle, exponential amplification of the specified fragment occurs.

In order to utilize this technology some DNA sequence information must be determined or known, either of the segment of interest or flanking regions, so that the oligonucleotide primers can be correctly synthesized. This can be a potential problem in some areas of research but is of no consequence when working with mitochondrial DNA since the entire genome has been sequenced (Anderson *et al.* 1981). After the target region has been amplified by the polymerase chain reaction, it is still necessary to analyze the DNA by another method such as direct DNA sequencing.

The power of this methodology was illustrated in a study by Higuchi *et al.* (1988) in which the polymerase chain reaction was used to examine polymorphic DNA sequences from single human hairs. The amplification products, a part of the D-loop region of mtDNA, were evaluated by direct DNA sequencing, hybridization with allele-specific oligonucleotide probes and the determination of amplified DNA fragment length differences. The PCR enabled the researchers to detect genetically variable mitochondrial and nuclear DNA sequences from the root region of shed, as well as freshly-plucked, single hairs.

Recently our laboratory developed a method to generate single stranded DNA for dideoxy sequencing following amplification by the Polymerase Chain Reaction (Mitchell and Merril 1989). This technique incorporates biotin into one of the amplification primers. The DNA strand produced by the biotinylated primer can be isolated from the homologous strand on a column containing streptavidin agarose. In this manner, the two DNA strands can be separated thereby producing a single stranded template for use in dideoxy sequencing protocols.

MITOCHONDRIAL DNA AND HUMAN DISEASES

Diseases characterized by disordered function of respiratory chain components encoded wholly or partly by mitochondrial DNA, or those with a maternal pattern of inheritance have been suspected of being the result of mutations in the mitochondrial genome. One such disease, Leber's optic neuropathy, is a hereditary cause of blindness which is exclusively maternally transmitted; it affects males more commonly than females but is clearly not X-linked. Until recently, the methodologies available to investigate these hypothesis have been the limiting step in discovering the true nature of these diseases. For example, Holt *et al.* (1988 c) published the results of restriction endonuclease analysis of leukocyte mitochondrial DNA in patients with Leber's optic atrophy and they concluded that there were no differences between affected and unaffected individuals from the same maternal line.

However, Wallace *et al.* (1988) and Singh *et al.* (1989) utilizing the more powerful DNA sequencing and the polymerase chain reaction methods, determined that a replacement mutation in the mtDNA sequence at nucleotide 11778 was associated with Leber's hereditary optic neuropathy in nine of the eleven large pedigrees studied. This particular mutation results in the conversion of a highly conserved arginine to a histidine at codon 340 in the NADH dehydrogenase subunit 4 gene. In these nine lineages, the researchers concluded that this mutation is necessary for the expression of the disease phenotype. Parker *et al.* (1989) postulated that the clinical variation of the disease could be explained on the basis of heteroplasmy (that is, the variable amounts of the abnormal mitochondrial genome initially present in the ovum). Since there were also two discordant pedigrees, there may be other mutations which could lead to related phenotypes. Knowing the casual mutation may eventually lead to an understanding of the pathophysiology of the symptoms which may, in turn, lead to clinical treatments.

Another group of diseases which are being investigated as to their relationship with mitochondrial DNA are the mitochondrial myopathies and encephalopathies. These are a clinically and biochemically heterogeneous group of neurological disorders characterized histologically by the ragged red fibers seen in muscle biopsy specimens (Harding 1989). This category includes encephalomyopathy, Kearns-Sayre syndrome, progressive external ophthalmoplegia (PEO) and various myopathies. Recent studies (Holt *et al.* 1988a; Saifuddin Noer *et al.* 1988; Zeviani *et al.* 1988; Lestienne and Ponsot 1988; Ozawa *et al.* 1988; Moraes *et al.* 1989; Schon *et al.* 1989) have found varying deletions in the mitochondrial DNA in a proportion of the patients with these clinical presentations. These deletions ranged on size from 0.4 to 7.6 Kb and were primarily observed only in skeletal muscle. One study (Saifuddin Noer *et al.* 1988), however, observed the deletion in skeletal muscle and in liver mtDNA. The presence of deletions was strongly but not absolutely correlated with certain clinical syndromes.

Duplications of mitochondrial DNA have also been documented (Poulton *et al.* 1989). This study described two patients with a distinctive clinical picture of mitochondrial myopathy, including diabetes mellitus, who were found to have abnormal mitochondrial genomes with a large duplication (about 8 Kb). In these individuals, the abnormal mtDNAs were confined to skeletal muscle and normal genomes were present in all tissues. The finding that there were no detectable duplicated mitochondrial molecules in the unaffected family members suggests that the association of the abnormal mtDNA with mitochondrial myopathy is causative. The investigators postulated that the variation in proportions of abnormal mtDNA and the differing energy demands of various tissues may explain the phenotype of individual patients.

Given the monoclonal, maternal inheritance pattern of the mitochondrial genome, its large number of copies, and its potential to provide individualization and information of maternal ethnic origin, this genome will prove its importance in future forensic studies. This will become increasingly apparent as the technologies to examine DNA sequences continue to improve.

REFERENCES

Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., *Staden, R. and Young, I. G. (1981).* Sequence and organization of the mitochondrial genome, Nature 290:457–465.

Aquadro, C. F. and Greenberg, B. D. (1983). Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals, Genetics 103:287–312.

Blanc, H., Chen, K. H., D'Amore, M. A. and Wallace, D. C. (1983). Amino acid change associated with the major polymorphic Hinc II site of Oriental and Caucasian mitochondrial DNAs, Am. J. Hum. Genet. 35:167–176.

Bonne-Tamir, B.B., Johnson, M.J., Natali, A., Wallace, D.C. and Cavalli-Sforza, L. L. (1986). Human mitochondrial DNA types in two Israeli populations. A comparative study at the DNA level, Am. J. Hum. Genet. 38:341–351.

Brega, A., Gardella, R., Semino, O., Morpurgo, G., Astaldi Ricotti, G. B., Wallace, D.C. and Santachirara-Benerecetti, A. S. (1986a). Genetic studies on the Tharu population of Nepal: restriction endonuclease polymorphisms of mitochondrial DNA, Am. J. Hum. Genet. 39:502–512.

Brega, A., Scozzari, R., Maccioni, L., Iodice, C., Wallace, D. C., Bianco, I., Cao, A. and Santachirara-Benerecetti, A. S. (1986b). Mitochondrial DNA polymorphisms in Italy. I. Population data from Sardinia and Rome, Am. J. Hum. Genet. 50:327–338.

Brown, W. M. (1980). Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis, Proc. Natl. Acad. Sci. USA 77:3605–3609.

Brown, W.M., George, M. Jr. and Wilson, A. C. (1979). Rapid evolution of animal mtDNA, Proc. Natl. Acad. Sci. USA 76:1967–1971.

Brown, W. M. and Goodman, H. M. (1979). Quantitation of intrapopulation variation by restriction endonuclease analysis of human mitochondrial DNA, In: Extrachromosomal DNA (Cummings, D. J., Borst, P., David, I. B., Weissman, S. M. and Fox, C. F., eds.). Academic Press, NY, pp. 485–499.

Brown, W. M., Prager, E. M., Wang, A. and Wilson, A. C. (1982). Mitochondrial DNA sequences of primates: tempo and mode of evolution, J. Mol. Evol. 18:225–239.

Cann, R. L., Brown, W. M. and Wilson, A. C. (1984). Polymorphic sites and the mechanism of evolution in human mitochondrial DNA, Genetics 106:479–499.

Cann, R. L., Stoneking, M. and Wilson, A. C. (1987). Mitochondrial DNA and human evolution, Nature 325:31–36.

Case, J. T. and Wallace, D. C. (1981). Maternal inheritance of mitochondrial DNA polymorphisms in cultured human fibroblasts, Somat. Cell. 7:103–108.

Cotton, R. G., Rodriguez, N. R. and Campbell, R. D. (1988). Reactivity of cytosine and thiamine in single-base-pair mismatches with hydroxylamine and osmiumtetroxide and its application to the study of mutations, Proc. Natl. Acad. Sci. USA 85:4397–4401.

Denaro, M., Blanc, H., Johnson, M. J., Chen, K. H., Wilmsen, E., Cavalli-Sforza, L. L. and Wallace, D. C. (1981). Ethnic variation in Hpa I endonuclease cleavage patterns of human mitochondrial DNA, Proc. Natl. Acad. Sci. USA 78:5768–5772.

Ferris, S., Wilson, A. C. and Brown, W. M. (1981). Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA, Proc. Natl. Acad. Sci. USA 78:2432–2436.

Fukunga, M. and Yielding, K. L. (1979). Fate during cell growth of yeast mitochondria and nuclear DNA after photolytic attachment of the monazide analog of ethidium, Biochem. Biophys. Res. Commun. 90:582–586.

Giles, R. E., Blanc, H., Cann, H. M. and Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA, Proc. Natl. Acad. Sci. USA 77:6715–6719.

Harding, A. E. (1989). The mitochondrial genome—breaking the magic circle, N. Eng. J. Med. 320:1341–1343.

Harihara, S., Hirai, M. and Omoto, K. (1986). Mitochondrial DNA polymorphism in Japanese living in Hokkaido, Jpn. J. Hum. Genet. 31:73–83.

Hausworth, W. W., Van de Walle, M. J., Lapis, P. J. and Olivo, P.D. (1984). Heterogenous mitochondrial DNA D-loop sequences in bovine tissue, Cell 37:1001–1007.

Higuchi, R., won Beroldingen, C. H., Sensabaugh, G. F. and Erlich, H. A. (1988). DNA typing from single hairs, Nature 332:543–546.

Holt, I. J., Cooper, J. M., Morgan-Hughes, J. A. and Harding. A. E. (1988a). Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies, Lancet i:1462.

Holt, I. J., Harding, A. E. and Morgan-Hughes, J. A. (1988b). Mitochondrial DNA polymorphism in mitochondrial myopathy, Nature 331:717–719.

Holt, I. J., Miller, D. H. and Harding, A. E. (1988c). Restriction endonuclease analysis of leukocyte mitochondrial DNA in Leber's optic neuropathy, J. Neurology 51:1075–1077.

Horai, S., Gojobori, T. and Matsunaga, E. (1984). Mitochondrial DNA polymorphisms in Japan. I: analysis with restriction enzymes of 6 base pair recognition, Hum. Genet. 68:324–332.

Horai, S. and Matsunaga, E. (1986). Mitochondrial DNA polymorphism in Japanese. II: analysis with restriction enzymes of four or five base pair recognition, Hum. Genet. 72:105–117.

Hutchinson, C. A. III, Newbold, J. E., Potter, S. S. and Edgell, M. H. (1974). Maternal inheritance of mammalian mitochondrial DNA, Nature 251:536–538.

Johns, D. R., Drachman, D. B. and Hurko, O. (1989). Identical mitochondrial DNA deletion in blood and muscle, Lancet i:393–394.

Kunkel, T. A. and Loeb, L. A. (1981). Fidelity of mammalian DNA polymerases, Science 213:765–767.

Lestienne, P. and Ponsot, G. (1988). Kearns-Sayre syndrome with muscle mitochondrial DNA deletion, Lancet i:885.

Li, H., Gyllensten, U. B., Cui, X., Saiki, R. S., Erlich, H. A. and Arnheim, N. (1988). Amplification and analysis of DNA sequences in single human sperm and diploid cells, Nature 335:414–417.

Maxam, A. M. and Gilbert, W. (1977). A new method for sequencing DNA, Proc. Natl. Acad. Sci. USA 74:560–564.

Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K. and Green, M. R. (1984). Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter, Nucleic Acids Res. 12:7035–7056.

Minishu, Y., Xinfang, Q., Jinglun, X., Zudong, L., Jiazhen, T., Houjun, L., Dexiang, L., Li, L., Wuzhong, Y., Xianzhen, T. and Meirong, Z. (1988). Mitochondrial DNA polymorphism in Chinese, Scientia Sinica 7:860–872.

Mitchell, L. G. and Merril, C. R. (1989). Affinity generation of single-stranded DNA for dideoxy sequencing following the polymerase chain reaction, Anal. Biochem. 178:239–242.

Miyaki, M., Yatagaai, K. and Ono, T. (1977). Strand breaks of mammalian mitochondrial DNA induced by carcinogens, Chem. Biol. Interactions 17:321–329.

Monnat, R. J. and Loeb, L. A. (1985). Nucleotide sequence preservation of human mitochondrial DNA, Proc. Natl. Acad. Sci. USA 82:2895–2899.

Monnat, R. J., Maxwell, C. L. and Loeb, L. A. (1985). Nucleotide sequence preservation of

human leukemic mitochondrial DNA, Cancer Res. 45:1809–1814.

Monnat, R. J. and Reay, D. T. (1986). Nucleotide sequence identity of mitochondrial DNA from different human tissues, Gene 43:205–211.

Moraes, C. T., Di Mauro, S., Zeviani, M., Lombes, A., Shanske, S., Miranda, A. F., Nakase, H., Bonilla, E., Werneck, L. C., Servidei, S., Nonaka, I., Koga, Y., Spiro, A. J., Brownell, A. K. W., Schmidt, B., Schotland, D. L., Zupanc, M., DeVivo, D. C., Schon, E. A. and Rowland, L. P. (1989). Mitochondrial DNA deletions in progressive external ophthalmoplegia and Kearns-Sayre syndrome, N. Eng. J. Med. 320:1293–1299.

Mullis, K. B. and Faloona, F. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. In: Methods in Enzymology, v. 155 (R. Wu, ed.), Academic Press, San Diego, CA, pp. 335–350.

Mullis, K. B., Faloona, F., Scharf, S., Saiki, R., Horn, G. and Erlich, H. A. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction, Cold Spring Harbor Symp. Quant. Biol. 51 (1):263–273.

Myers, R. M., Larin, Z. and Maniatis, T. (1985). Detection of single base substitutions by ribonuclease cleavage at mismatches in RNA:DNA duplexes, Science 230:1242–1246.

Ozawa, T., Yoneda, M., Tanaka, M., Ohno, K., Sato, W., Suzuki, H., Nishikimi, M., Yamamoto, M., Nonaka, I. and Horai, S. (1988). Maternal inheritance of deleted mitochondrial DNA in a family with mitochondrial myopathy, Biochem. Biophys. Res. Commun. 154:1240–1247.

Parker, W. D., Oley, C. A. and Parks, J. K. (1989). A defect in mitochondrial electron-transport activity (NADH-coenzyme Q oxidoreductase) in Leber's hereditary optic neuropathy, N. Eng. J. Med. 320:1331–1333.

Potter, S. S., Newbold, J. E., Hutchison, C. A. and Edgell, M. H. (1975). Specific cleavage analysis of mammalian mitochondrial DNA, Proc. Natl. Acad. Sci. USA 72:4496–4500.

Poulton, J., Deadman, M. E. and Gardiner, R. M. (1989). Duplications of mitochondrial DNA in mitochondrial myopathy, Lancet i:236–240.

Robin, E. D. and Wong R. (1988). Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells, J. Cell. Physiol. 136:507–513.

Saifuddin Noer, A., Marzuka, S., Trounce, I. and Byrne, E. (1988). Mitochondrial DNA deletion in encephalomyopathy, Lancet ii:1253–154.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnheim, N. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia, Science 230:1350–1354.

Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1986). Analysis of enzymatically amplified beta-globin and HLA-DQalpha DNA with allele specific oligonucleotide probes, Nature 324:163–166.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988). Primer directed enzymatic amplification of DNA with thermostable DNA polymerase, Science 239:487–491.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. USA 74:5463–5468.

Santachiara-Benerecett, A. S., Scozzari, R., Semino, O., Torroni, A., Brega, A. and Wallace, D. C. (1988). Mitochondrial DNA polymorphisms in Italy. II: molecular analysis of new and rare morphs from Sardinia and Rome, Am. J. Hum. Genet. 52:39–55.

Schon, E. A., Rizzuto, R., Moraes, C. T., Nakase, H., Zeviani, M. and DiMauro, S. (1989). A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA, Science 244:346–349.

Scozzari, R., Torroni, A., Semino, O., Sirugo, G., Brega, A. and Santachiara-Benerecetti, A. S. (1988). Genetic studies on the Senegal population. I. Mitochondrial DNA polymorphisms, Am. J. Hum. Genet. 43:534–544.

Singh, G., Lott, M. T. and Wallace, D. C. (1989). A mitochondrial DNA mutation as a cause of Leber's hereditary optic neuropathy, N. Eng. J. Med. 320:1300–1305.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis, J. Mol. Biol. 98:503–517.

Wallace, D. C., Singh, G., Lott, M. T., Hodge, J. A., Schurr, T. G., Lezza, A. M. S., Elsas, L. J. II and Nikoskelainen, E. K. (1988). Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy, Science 242:1427–1430.

Zeviani, M., Moraes, C. T., DiMauro, S., Nakase, H., Bonilla, E., Schon, E. A. and Rowland, L. P. (1988). Deletions of mitochondrial DNA in Kearns-Sayre syndrome, Neurology 38:1339–1346.

APPENDIX

ADVANTAGES OF THE MITOCHONDRIAL GENOME

1. There are many copies, 10^{17}
2. It is robust.
   Svante Paabo [PNAS, *86:* 1939–1943, 1989] was able to amplify and study short mitochondrial sequences from:
   13,000 year old skin specimen from ground sloth, Chile
   5,000 year old skin from a natural mummy, Egypt
3. Genome can provide ethnic or geographic origin information.
4. Genome can provide information on individualization.

DISADVANTAGES OF THE MITOCHONDRIAL GENOME

1. It may not be monoclonal in some individuals.
2. Ethnic or geographic origin information is maternal.
3. Cannot differentiate between family members within common maternal line.
4. May be of limited value in inbred communities.

TASKS REMAINING TO DEVELOP MIT-GENOME INTO A FORENSIC TOOL

1. Somatic mutation studies in normal and diseased individuals, with both mitotic and postmitotic tissues.
2. Family studies to determine how many generations in the human maternal line are required to differentiate individuals.
3. Ethnic or geographic origin studies to refine the specific patterns that are present and to determine their reliability in the populations of interest.
4. Technical refinements in the automation of the PCR amplification and sequencing.

A STATISTICAL APPROACH FOR VNTR ANALYSIS

Bruce Budowle and Keith L. Monson

FBI Laboratory
Quantico, Virginia

The tools of the molecular biologist that enable the detection of restriction fragment length polymorphisms (RFLPs) are having a major impact on the field of forensic science. By using a panel of DNA probes, each of which recognizes an independent hypervariable locus, sufficient data can be provided to produce a composite DNA profile which is unique to an individual (excluding monozygotic twins). In fact, these hypervariable DNA regions, which are comprised of a variable number of tandem repeats (VNTRs), (Wyman and White 1980; Nakamura *et al.* 1987ab; Wong *et al.* 1987), currently provide the forensic scientist the best avenue to exclude an individual who has been falsely associated with an evidentiary sample.

For the situations where attempts to exclude a suspect as being a possible contributor of a sample via VNTR typing prove fruitless, a statistical value is needed to be placed upon the DNA profile to determine the portion of the population that could be potential contributors of the evidentiary sample(s). Thus, proper weight can be applied to matching results. A statistical approach must be developed that takes into consideration limitations of the technology and available population data bases so that undue weight will not be placed on a sample attributed to an accused individual.

RESOLUTION

The solution of alleles that differ by one to a few repeat sequences is not possible by present electrophoretic technology. This is particularly so when the overall size of a DNA fragment (or allele) is large and the core repeat sequence is short. Furthermore, it has never been demonstrated that VNTR alleles vary in size by increments of the core sequence length. It is entirely possible that alleles could differ in size by as little as one nucleotide. Therefore, as expected, the measurements of alleles from a sample population form a quasi-continuous distribution of allele sizes. Until resolution of one nucleotide differences in size can be achieved, the data cannot be reduced to discrete alleles. Analysis of a quasi-continuous distribution of alleles differs significantly from analyses of the traditional genetic marker systems which provide discrete allele data.

The resolution issue is further complicated by the fact that the resolving capability of the electrophoretic system changes continuously across the gel length. This phenomenon is not easily addressed mathematically for matching alleles from two different specimens and may have to be calculated for each electrophoretic run (Grossman, L., personal communication). Additionally, resolution is related to width of the RFLP bands. With isotopic labels and film-based detection, band width is an uncontrollable value that will vary due to the quantity of DNA and the autoradiographic exposure time. Therefore, a value related to the ability to resolve two band fragments based on the physical space between them should be avoided.

SAMPLE POPULATIONS

At present, there are few data on the distribution of VNTR alleles for particular loci for various racial and ethnic groups. Therefore, there is no evidence to support the assertion that a sample population adequately represents the true population or other subpopulation groups. For other polymorphic genetic markers, such as HLA, differences are observed in the frequency of certain alleles among different ethnic groups. For example, the frequencies of HLA–A1 in Japanese and Chinese are 0.2% and 2.3%, respectively (Bauer *et al.* 1984). Therefore, if a Japanese sample population was used for a data base to represent all Orientals, undue weight would be placed on a sample attributed to a Chinese individual carrying the HLA–A1 allele. The same situation can and probably will hold true for VNTR genetic markers and racial subgroups.

Although it has been suggested at least 150 individuals are required for a sample population (Odelberg *et al.* 1989), the actual minimum number is not easily addressed. The number will depend upon the particular statistical approach and the polymorphic nature of the VNTR locus. For the D1S7 locus it has been reported that the core repeat size is nine nucleotides long (Jeffreys *et al.* 1988). Sizing data on this locus (generated by the FBI) suggest that alleles range in size from less than 1000 base pairs to approximately 20,000 base pairs. Therefore, there is the potential for more

than 2000 different alleles to exist. It is difficult to imagine that a sample population of 150–250 people would adequately represent the true frequency of some of the alleles of the population (particularly the low frequency alleles). Thus, sampling error alone could contribute to an underestimation of the frequency of occurrence of a particular allele carried by an individual.

An alternative example may illustrate the issue better. It is known that the odds of obtaining a heads or tails result from the toss of a properly weighted coin is 0.5. Given a sufficient number of tosses the observed frequency of occurrence of heads or tails should be close to 0.5. However, if only ten tosses were made it is entirely possible, for example, that three heads and seven tails could be observed (a deviation by chance from the true value). If this small number of tosses was used to represent the true frequency of heads and tails in a coin toss population, some values would be underestimates of the true value. For instance, the frequency of occurrence of a head-head combination for two simultaneous coin tosses would be estimated as 0.09 instead of the true value of 0.25.

BINNING

A statistical approach is needed that appropriately addresses the aforementioned limitations of technology, the genetics of VNTRs and available data base. The statistical application should not exceed the available data and should be utilized with the premise that placing undue weight (underestimation of frequency of occurrence) on the evidence attributed to an accused individual is unacceptable.

Arbitrarily defined bins were created to accommodate continuous data so that alleles could be statistically classified, the number of samples required for a sample population could be reduced, and an overestimate of an allele's frequency would be obtained to compensate for differences that might occur in racial subgroups or due to sampling error.

The boundaries of the bins were determined using size standard markers (Table 1). The fragments that comprise the size standards are derived from restriction digests of viral DNAs. Since the particular viral DNAs have been completely sequenced, the exact base pair length of each fragment (and thus the bin boundary dimensions) is known. The only requirement for the boundaries of each bin is that the difference in size of the two fragments that fix the boundaries of a bin must be

greater than the determined measurement imprecision of the analytical system.

Table 1. SIZE (IN BASE PAIRS) OF VIRAL DIGESTS [a] FOR SIZE STANDARDS

12,830	3,980	1,789
11,369	3,675	1,638
10,094	3,330	1,508
8,453	3,034	1,353
7,242	2,863	1,197
6,369	2,693	1,078
5,686	2,523	964
5,220	2,352	872
4,822	2,089	773
4,324	1,925	640

[a] Size standards obtained from the Lifecodes Corporation (Valhalla, NY).

Based upon size, alleles from a sample population can be placed or clustered within an appropriate bin. The bins are sufficiently large so that more than one type of allele can reside within a bin. The total number of alleles from a sample population that resides within a bin are used to calculate a frequency of occurrence for that bin. Then, an allele from an unknown sample that falls within a particular bin will be assigned that bin's frequency for statistical evaluation. When an allele falls on the boundary of two bins (that is, overlaps two bins after adjustment for measurement imprecision or its band width is such that it physically overlaps more than one bin), the bin with the largest frequency is selected for statistical purposes. Thus, the estimate of the frequency of occurrence of an allele or set of alleles from an individual (or questioned sample) will be greater than the true frequency of alleles in the sample population. It should be noted that the binning process has nothing to do with matching criteria. It is entirely possible that two different alleles can reside within the same bin and not constitute a match. The two different alleles, however, will be assigned the same allelic frequency.

An example of Black population data for the D14S13 locus (Nakamura *et al.* 1988) using the binning approach is shown in Table 2. Using the size markers listed in Table 1, 31 bins were created. All but six of the bins contain observed events. Table 3 provides Black population data for the D17S79 locus (Balasz *et al.* 1989). Of the 31 possible bins, only 16 contained observable events. A comparison of Tables 2 and 3 provides a rapid manner for evaluating the polymorphic nature of each of the VNTR systems. It is clearly evident that D14S13 is more informative for genetic characterization of evidentiary samples than D17S79 (although the

more important criteria of sensitivity of detection favors the probe that hybridizes to D17S79).

Table 2. BINNED BLACK POPULATION DATA FOR D14S13

Bin	Range (bp)	Allele Count	95% LCL [a,c]	Point Est.	95% UCL [b,c]
1	0–639	0	.000	.000	.000
2	640–722	0	.000	.000	.000
3	773–871	7	.006	.017	.051
4	872–963	9	.008	.022	.058
5	964–1077	7	.006	.017	.051
6	1078–1196	5	.003	.012	.044
7	1197–1352	22	.029	.055	.101
8	1353–1507	28	.039	.069	.119
9	1508–1637	20	.025	.050	.095
10	1638–1788	33	.049	.082	.134
11	1789–1924	18	.022	.045	.088
12	1925–2088	31	.045	.077	.128
13	2089–2351	37	.056	.092	.146
14	2352–2522	27	.038	.067	.116
15	2523–2692	23	.031	.057	.104
16	2693–2862	12	.013	.030	.069
17	2863–3033	17	.020	.042	.085
18	3034–3329	14	.016	.035	.075
19	3330–3674	36	.054	.089	.143
20	3675–3979	7	.006	.017	.051
21	3980–4323	26	.036	.065	.113
22	4324–4821	7	.006	.017	.051
23	4822–5219	3	.001	.007	.036
24	5220–5685	1	.000	.002	.028
25	5686–6368	2	.001	.005	.032
26	6369–7241	9	.008	.022	.058
27	7242–8452	2	.001	.005	.032
28	8453–10093	0	.000	.000	.000
29	10094–11368	0	.000	.000	.000
30	11369–12829	0	.000	.000	.000
31	12830–	0	.000	.000	.000

[a] LCL = Lower Confidence Limit.
[b] UCL = Upper Confidence Limit.
[c] LCL and UCL calculated according to Quesenberry and Hurst (1964).

Table 3. BINNED BLACK POPULATION DATA FOR D71S79

Bin	Range (bp)	Allele Count	95% LCL [a,c]	Point Est.	95% UCL [b,c]
1	0–639	0	.000	.000	.000
2	640–772	1	.000	.002	.026
3	773–871	1	.000	.002	.026
4	872–963	0	.000	.000	.000
5	964–1077	11	.010	.026	.061
6	1078–1196	20	.024	.047	.089
7	1197–1352	122	.222	.284	.355
8	1353–1507	112	.201	.260	.330
9	1508–1637	46	.069	.107	.162
10	1638–1788	23	.029	.053	.098
11	1789–1924	21	.025	.049	.092
12	1925–2088	28	.037	.065	.112
13	2089–2351	18	.021	.042	.083
14	2352–2522	7	.005	.016	.048
15	2523–2692	4	.002	.009	.038
16	2693–2862	3	.001	.007	.034
17	2863–3033	0	.000	.000	.000

Table 3. BINNED BLACK POPULATION DATA FOR D71S79—Continued

Bin	Range (bp)	Allele Count	95% LCL [a,c]	Point Est.	95% UCL [b,c]
18	3034–3329	6	.004	.014	.045
19	3330–3674	7	.005	.016	.048
20	3675–3979	0	.000	.000	.000
21	3980–4323	0	.000	.000	.000
22	4324–4821	0	.000	.000	.000
23	4822–5219	0	.000	.000	.000
24	5220–5685	0	.000	.000	.000
25	5686–6368	0	.000	.000	.000
26	6369–7241	0	.000	.000	.000
27	7242–8452	0	.000	.000	.000
28	8453–10093	0	.000	.000	.000
29	10094–11368	0	.000	.000	.000
30	11369–12829	0	.000	.000	.000
31	12830–	0	.000	.000	.000

[a] LCL = Lower Confidence Limit.
[b] UCL = Upper Confidence Limit.
[c] LCL and UCL calculated according to Quesenberry and Hurst (1964).

Although binning of alleles is conservative and appears to compensate for fluctuations both between subgroups and due to sampling error, it does not address the low-frequency allele situation. For D14S13 bins 1, 2, 27–31 have no observable events and bins 24, 25 and 27 have only one or two events. It is these low frequencies that provide the numbers in the one in hundreds of millions to trillions range, and it is these low frequency values that there is the least confidence statistically (due to sampling error and subpopulation differences). An estimation of the upper confidence limit of a frequency of a bin with no events approaches two per cent (Lander, E., personal communication). Therefore, without adequate population data to support the assertion that these low frequency alleles truly represent the population, use of these values should be avoided. For a chi-square analysis, a minimum of five events is desired for data evaluation. In a similar fashion, the low frequency bins can be pooled to provide a minimum value of five alleles per bin. The number of bins for D14S13 Black population data is then 22 and the lowest number of observable events in a bin is 5 (Table 4). In fact, the confidence intervals would suggest that the point estimates are reasonable (particularly the low frequency bins) and thus the sample population size is adequate. Large frequency bins fluctuate slightly for different sample population subgroups but it is expected that there will be no placing of undue weight on a sample carrying a particular DNA profile (especially since the values already are extremely conservative).

123

Table 4. REBINNED [a] BLACK POPULATON DATA FOR D14S13

Bin	Range (bp)	Allele Count	95% LCL [b,d]	Point Est.	95% UCL [c,d]
1	0–871	7	.006	.018	.052
2	872–963	9	.008	.023	.059
3	964–1077	7	.006	.018	.052
4	1078–1196	5	.003	.013	.044
5	1197–1352	22	.029	.055	.102
6	1353–1507	27	.038	.068	.117
7	1508–1637	20	.026	.050	.095
8	1638–1788	32	.047	.080	.132
9	1789–1924	18	.022	.045	.089
10	1925–2088	31	.045	.078	.129
11	2089–2351	36	.055	.090	.144
12	2352–2522	27	.038	.068	.117
13	2523–2692	23	.031	.058	.105
14	2693–2862	12	.013	.030	.069
15	2863–3033	17	.021	.043	.086
16	3034–3329	14	.016	.035	.076
17	3330–3674	36	.055	.090	.144
18	3675–3979	7	.006	.018	.052
19	3980–4323	26	.036	.065	.114
20	4324–4821	7	.006	.018	.052
21	4822–6368	6	.005	.014	.048
22	6369–				
		11	.011	.028	.066

[a] N. B. one individual displayed a three-band pattern which is not included in the above tabulation.
[a] Low frequency bins must have a minimum of five events.
[b] LCL = Lower Confidence Limit.
[c] UCL = Upper Confidence Limit.
[d] LCL and UCL calculated according to Quesenberry and Hurst (1964).

Thus, the binning approach provides a statistical application for analysis of highly polymorphic, continuous data that reduces the number of samples required for a population data base, provides confidence that the result is not an underestimation of the frequency of occurrence of a set of alleles so undue weight is not placed on a sample attributed to an accused individual, and offers a portable data base system that can be transferred easily to any laboratory performing the analysis.

A fixed bin approach was chosen instead of floating bins (Balasz *et al.* 1989), again to avoid sensitivity to statistical bias. For a floating bin approach, the size value assigned to an allele in an evidence sample is assumed to be the mean value. This may result in assessing a low frequency portion of the allele data distribution for a particular population sample when it would be inappropriate. At present, it cannot be demonstrated that the size value placed on an allele is the mean value (and often it may not be the mean value), and therefore, a floating bin approach was avoided. However, when approaches are implemented that will permit the mean value to be determined (such as the use of invariant bands in the sample), the floating bin approach could be reconsidered.

HARDY-WEINBERG EQUILIBRIUM

The Hardy-Weinberg equilibrium is based on the simple principle of probability that the chance of two independent events (that is, alleles) occurring simultaneously is the product of their probabilities times the number of different combinations of the two independent events (Hardy 1908). Thus, the Hardy-Weinberg rule provides a simple algebraic approach for estimating gene frequencies in populations. The rule assumes random selection or independence of alleles at a particular locus. Factors such as mutation rate, natural selection, social customs, and ethnic and religious stratifications may place some restriction on randomness. However, except for a few cases such as sickle-cell anemia (where one of the genotypes has reduced viability), these factors do not produce detectable deviations in the Hardy-Weinberg equilibrium for one sampling of a population.

The rule is often referred to as the Hardy-Weinberg law; but the word law may give more prominence to the rule than should be placed upon it. Therefore, some scientists may attempt to apply the conventional formulation of the Hardy-Weinberg rule when inappropriate. The conventional formulation of the Hardy-Weinberg rule, requires discrete alleles and that there be no measurement imprecision. Neither of these requirements exist for VNTR loci that are analyzed by the standard RFLP technique. The applicability of the Hardy-Weinberg rule is exacerbated further since some VNTR homozygotes may actually be heterozygotes. Thus, the number of observed homozygotes will almost always be greater than the number of expected homozygotes (thus, the Wahlund effect (Wahlund 1928) would not apply either). Table 5 shows the data regarding observed and expected homozygotes for a number of VNTR loci and as predicted (by these authors) the number of observed homozygotes is always greater than expected (even for those loci in equilibrium—see Table 5). These authors were able to predict *a priori* which VNTR systems would be in equilibrium based solely on the base pair size of the alleles and the sensitivity of detection of the probe. Since larger DNA fragments (which contain more repeat sequences) provide a more intense signal than smaller fragments, it is entirely possible that some small sized alleles go undetected. Thus, it is reasonable to predict that null alleles must exist. Recent data on D2S44 demonstrate that this is in fact the case (Eisenberg, A., personal communication). Similar observations have been found for D16S85

(Allen, personal communication). He observed that some DNAs digested with Pvu II that were heterozygotes for the D16S85 locus became single band patterns when digested with Hae III (due to the generation of smaller fragments by Hae III). Therefore, there should always be more observed homozygotes than expected, and as the population data base size increases even the VNTR loci (D2S44, D14S13, and D4S139) apparently in statistical equilibrium (in Table 5) will eventually fall out of equilibrium. A null-null individual for a VNTR locus has yet to be observed (and would be difficult to verify without family data) due to the limited size of the data base, but expectations are that these individuals will eventually be observed.

Table 5. DATA REGARDING OBSERVED AND EXPECTED HOMOZYGOTES FOR VNTR LOCI

Locus Race [a]		Homozygotes [b]			Chi-Square [c] Equilibrium
		Observed	Expected		
D2S44	C	20	17.49	0.36	yes
D14S13	C	21	19.32	0.15	yes
D4S139	C	18	14.26	0.98	yes
D17S79	C	63	43.07	9.22	no
D16S85	C	27	17.56	5.07	no
D1S7	C	20	12.07	5.21	no
D2S44	B	20	14.71	1.90	yes
D14S13	B	13	12.06	0.07	yes
D4S139	B	19	18.02	0.05	yes
D17S79	B	52	37.61	5.51	no
D16S85	B	35	32.18	0.25	yes
D1S7	B	19	12.33	3.61	yes

[a] C = Caucasian, B = Black.
[b] For this analysis homozygotes were considered either single band patterns or heterozygotes where both bands reside within one bin.
[c] Degrees of freedom = 1.

The existence of these null alleles, continuous data, and measurement imprecision make the Hardy-Weinberg rule inappropriate for addressing the genetic make-up of the sample population. There may be some restriction on randomness for these VNTR loci, but it is true that for the vast majority of inherited characteristics the alleles at each locus combine essentially at random. Further, the odds of discovering so many loci that are affected by some form of selection seem unreasonable. Instead, a reasonable assumption can be made. People are unaware of their VNTR genetic composition and the VNTR genotype does not enter into the decision to have offspring. Therefore, the possibility of social stratification, mentioned earlier, can be neglected and the alleles at a locus can be considered independent of each other. Therefore, the algebraic approaches put forth in the Hardy-Weinberg rule can be applied. The final empirical value should be a reasonable estimate of the frequency of occurrence in the sample population of the particular set of alleles.

MATCHING CRITERIA

Now that a statistical approach is available, the measurement imprecision of allele sizes must be determined. Band width, electrophoretic resolution, and slight variation in electrophoretic mobility of DNA fragments can affect measurement imprecision. This value should be determined empirically by repetitive measurements. It is not the ability to resolve two fragments as different under ideal conditions but rather the latitude in measurements of the same sample that must be considered for matching criteria. The data should be derived from ideal samples (that is, controlled DNA quantities and autoradiographic exposure times) as well as nonideal samples (that is, varied DNA quantities and autoradiographic exposure times). Special attention should be paid to victim's DNA on the substrata to determine the measurement imprecision for forensic samples. The data for the FBI would suggest that size measurements of some fragments can vary +2.5% (or approximately a total of 5%) difference can exist (Monson and Budowle 1990). To date this maximum value of measurement imprecision has never been used by an FBI examiner to corroborate a visual match between two samples. In fact differences more than +2% have been categorized inconclusive.

When determining identity between two samples, a visual match is made first and it is confirmed (in the FBI Laboratory) by computer assisted image analysis after compensation for measure-

ment imprecision. These measurement imprecision values are expected to decrease in the near future with the implementation of an analysis that detects invariant (or monomorphic) bands. Fragments of DNA, common in size to almost all individuals, can serve as internal standards, should permit the demonstration of true band shifts, and reduce the questions of measurement imprecision.

CONCLUSION

The VNTR probe-locus systems presently are the most appropriate for characterization of body fluid/tissue specimens. They are highly polymorphic and the sensitivity of detection is such that many samples with limited quantities of DNA can still be analyzed. A statistical approach is available which compensates for sampling error, for differences in racial subgroups, limited sample population size, and limited RFLP resolution. Using the binning approach, it is unlikely that an array of alleles in a sample will be assessed a frequency of occurrence that is lower than the true frequency in the population (for unrelated individuals).

REFERENCES

Balasz, I., Baird, M., Clyne, M. and Meade, E. (1989). Human population genetic studies of five hypervariable DNA loci, Amer. J. Hum. Genet. 44:182–190.

Baur, M. P., Neugebauer, M., Deppe, H., Sigmund, M., Luton, T., Mayr, W. R. and Albert E. D. (1984). Population analysis on the basis of deduced haplotypes from random families. In: Histocompatibility Testing 1984, (Albert, E. D., Baur, M. P., and Mayr, W. R. eds.), Springer-Verlag, Berlin, pp. 333–341.

Jeffreys, A. J., Royle, N. J., Wilson, V. and Wong, Z. (1988). Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA, Nature 323:278–281.

Hardy, G. H. (1908). Mendelian proportions in a mixed population, Science 28:49–50.

Monson, K. L. and Budowle, B. (1990). A system for semi-automated analysis of DNA autoradiograms, In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C., pp. 127–132.

Nakamura, Y. Leppert, M., O'Connell, P., Wolff, R. Holm, T. Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987a). Variable number of tandem repeat markers for human gene mapping, Science 235:1616–1622.

Nakamura, Y., Gillilan, S., O'Connell, P., Leppert, M., Lathrop, G. M., Lalouel, J. M. and White, R. (1987b). Isolation and mapping of a polymorphic DNA sequence pYNH24 on chromosome 2 (D2S44), Nuc. Acids Res. 15:10073.

Nakamura, Y., Culver, M., Gill, J., O'Connell, P., Leppert, M., Lathrop, G., Lalouel, J. M. and White, R. (1988). Isolation and mapping of a polymorphic DNA sequence pMLJ14 on chromosome 14 (D14S13), Nuc. Acids Res. 16:381.

Odelberg, S. J., Plaetke, R., Eldridge, J. R., Ballard, P., O'Connell, P., Nakamura, Y., Leppert, M., Lalouel, J. M. and White, R. (1989). Characterization of eight VNTR loci by agarose gel electrophoresis, Genomics 5:915–924.

Quesenberry, C. P. and Hurst, D. C. (1964). Large sample simultaneous confidence intervals for multinomial proportions, Technometrics 6:191–195.

Wahlund, S. (1928). Zusammensetzung von populationeng und Korrelationserscheinungen von standpunkt der Vererbungslehre aus betrachet, Hereditas 11:65–106.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. J. (1987). Characterization of a panel of highly variable ministatellites cloned from human DNA, Ann. Hum. Genet. 51:264–288.

Wyman, A. R. and White, R. (1980). A highly polymorphic locus in human DNA, Proc. Natl. Acad. Sci. USA 77:6754-6758.

A SYSTEM FOR SEMI-AUTOMATED ANALYSIS OF DNA AUTORADIOGRAMS

Keith L. Monson and Bruce Budowle

FBI Laboratory
Quantico, Virginia

Routinely, a forensic serology laboratory is asked to determine whether or not a body fluid stain recovered during a crime scene search might have originated from a suspect in custody. Despite the advances in forensic serology in recent years, DNA technology potentially offers enhanced discrimination of individuals as well as significantly higher detection sensitivity to samples of limited quantity. Polymorphisms of DNA will not vary throughout the life of an individual, and thus are very useful for the purpose of discriminating one person from another. To detect the highly polymorphic genetic information available in human variable number tandem repeat sequences (VNTR's), DNA is subjected to restriction fragment length polymorphism (RFLP) analysis. The particular fractionated VNTR sequences are hybridized with a radioactively labeled probe and subsequently visualized on an autoradiogram (Budowle 1990). Since genetic polymorphism of VNTR's is based on size differences, data on fragment lengths are desirable. In order to rapidly and objectively determine fragment lengths an interactive image analysis system was developed (Monson 1988).

Fragment lengths (in base pairs, or molecular weight) can be determined by the distance traveled on the gel relative to DNA markers of known size. These molecular weight size markers are run in several lanes on each gel, always flanking the unknown sample (Figure 1). Fragment lengths are expressed in base pairs, and the numerical values resulting from specific VNTR polymorphism profiles with several probes can be used to confirm or refute a putative visual match of evidentiary samples with a suspect or to construct a profile which may be compared with those in a data base of convicted offenders or open case files.

There were several reasons an automated method of analyzing DNA autoradiograms was required. First, due to the power of the VNTR–RFLP method, a high case load was anticipated, necessitating rapid analysis. Secondly, since RFLP probes used are selected because they reveal so many phenotypes, closely-spaced bands can often appear on the autoradiogram. This can make accurate manual location difficult. Finally, it is important to have an objective method for fragment size
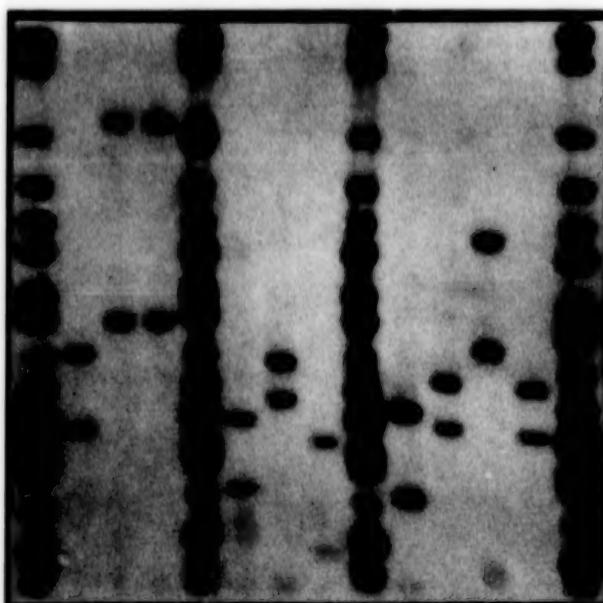

Figure 1. A typical autoradiogram produced by an RFLP probe for locus D2S44. The specimens in lanes 3 and 11 were duplicated in lanes 4 and 13, respectively. Molecular weight size markers are in lanes 1, 5, 9, and 14.

measurement, sufficiently precise to complement the scientist's visual evaluation of whether or not the DNA profile from two specimens are indistinguishable, and to support the eventual development of DNA profile data bases.

The software developed at the FBI for analysis of DNA autoradiograms is designed around readily available, off-the-shelf hardware. An IBM AT-compatible personal computer, outfitted with an image processing board, is the heart of the system (Figure 2). Approximate costs of duplicating the system range from $6–10,000, depending on the computer chosen to support the system (Monson 1988). This paper will describe briefly the operation of the system and discuss the use of molecular weight size markers.

THE FBI's ANALYTICAL SYSTEM

Semi-automated analysis of DNA autoradiograms using the FBI system comprises several operations:

1. Digitize the autoradiogram
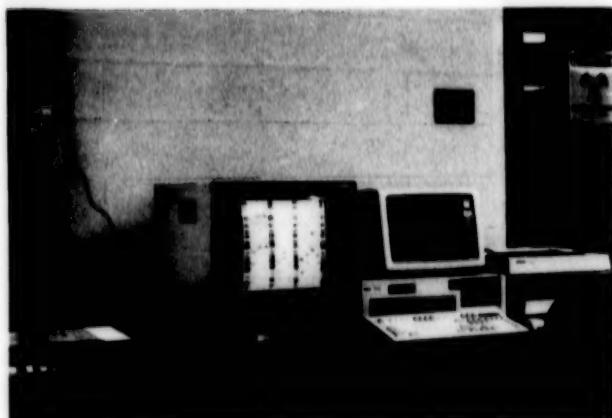
Figure 2. The FBI DNA image analysis system, which is based on an IBM PC/AT compatible microcomputer.

2. Locate lane boundaries
3. Determine band locations in size ladder
4. Perform geometric corrections
5. Determine band locations in specimen
6. Calculate fragment lengths

Each will be described in the following paragraphs.

Image Digitization

A digital version of the autoradiogram must be produced before calculations from the image can be made. A video camera mounted on a copy stand is used for image input, and a light box provides uniform illumination of autoradiograms (Figure 2). After the operator aligns an autoradiogram on the monitor, the image processing board digitizes the video signal and captures the image in computer memory. The image is then available for computer analysis. Digital images of 512 by 512 pixels are produced, with 256 levels of gray. This level of digital resolution is sufficient for the requirements of RFLP analysis. Different size gels are accommodated by simply changing the camera position.

Location of Lane Boundaries

After image digitization the lane boundaries must be located. The boundaries define the areas to be analyzed within the image. The operator specifies the number of bands present, and a feature location algorithm (Pun *et al.* 1985) automatically marks the likely boundaries (Figure 3). Lane boundaries are delineated by a vertical red line, nondestructive to the image.

The computer-generated boundaries can be modified if necessary. Since the entire region defined by the lane boundaries is used in subsequent steps to locate band positions, there is provision to
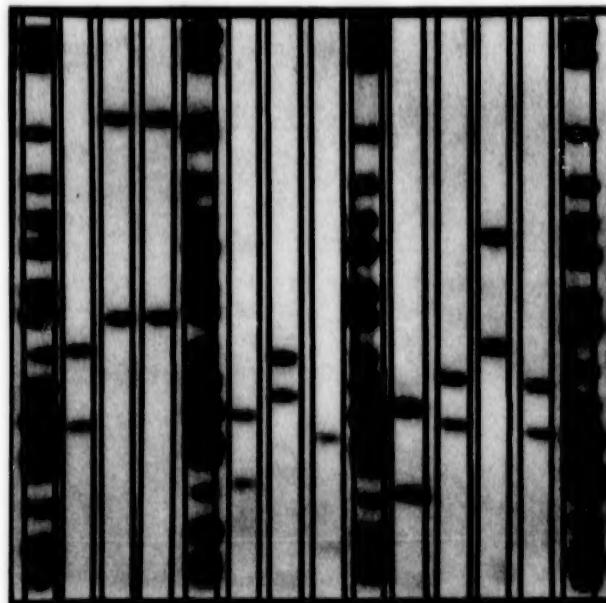

Figure 3. Computer-generated lane boundaries for the autoradiogram of Figure 1. Lane boundaries and lane numbers appear as color graphic overlays on the RBG image monitor.

modify the width of lane boundaries, if necessary to compensate for convex bands, curved lanes or autoradiogram noise. Lanes can also be added or deleted interactively.
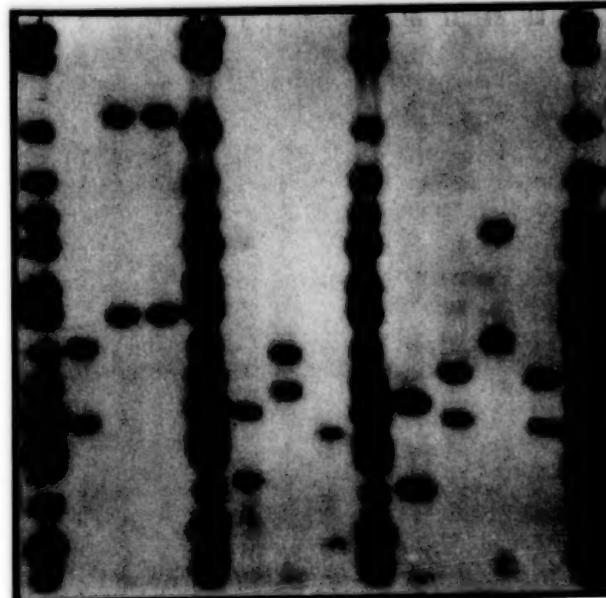

Figure 4. Computer-generated band locations for the molecular weight standards ladder in lane one.

Location of Standard Bands

The length of each band in the size marker ladder is known exactly. The ladder is produced by

multiple enzymatic digestions of viral DNAs, which have been fully sequenced. The locations of bands in the lanes containing size markers are used to correct for any inconsistencies in electrophoretic mobility attributable to any non-uniformities in electric field strength (following section). This information, along with the positions of specimen bands, allows the lengths of DNA fragments from the specimen to be calculated. Band locations are determined from densitometric traces calculated for each lane in the autoradiogram. Line scans for every column within each lane are summed to produce these densitometric traces. This increases statistical accuracy and minimizes the effects of lane curvature. The computer determines band locations by ranking local maxima in the density histograms. The number of bands expected in the size ladder and the exact molecular weight of each fragment are maintained in computer files and are confirmed by the operator during program initiation.
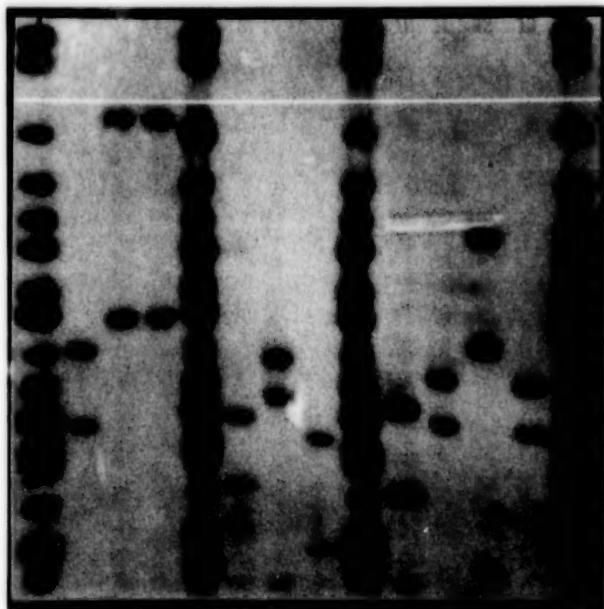

Figure 5. Interactive cursor used to disqualify a band mistakenly identified in lane one.

The operator must verify that the bands in each lane have been located correctly by the computer (Figure 4). When a band is mistakenly identified, a cursor is used to mark it for disqualification (Figure 5). There are three options for band replacement after disqualification of a band: (1) The automatic mode selects the next band from the list of candidates, ranked by band intensity. The candidates are determined by analyzing the density histogram for the lane. (2) Assisted manual relocation is most frequently used. The operator marks

the vicinity of the desired band (Figure 6), and the computer objectively locates its exact position. (3) The fully manual mode is occasionally used when automatic location is unsatisfactory, such as might occur in the case of severely overexposed bands.
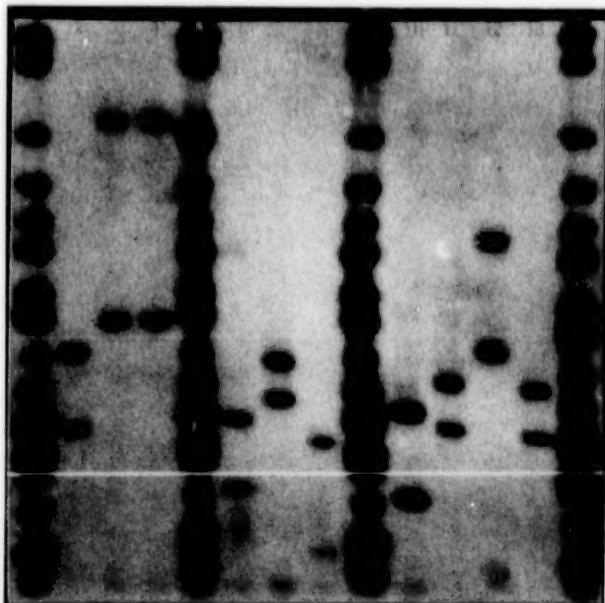

Figure 6. Cursor used to specify the replacement for the band in lane one which was marked for disqualification in Figure 5.

Geometric Correction

The lanes containing properly-marked standards (Figure 7) are used to correct for variations in electrophoretic mobility in each specimen lane across the gel. If the same specimen were run in every lane, corresponding bands would probably not align perfectly. To illustrate, duplicates of the specimens in lanes 3 and 11 have been placed in lanes 4 and 13, respectively, yet the bands do not align horizontally (Figure 1). Correction for non-uniform electrophoretic migration is made by first fitting lines between corresponding bands in lanes which contain standards. Lanes containing standards must flank the specimen lanes (Figure 1). An estimate of where the standards would have occurred in each specimen lane can be made by interpolation (the condition that the size ladder spans the expected range of fragment lengths is implicit). Thus, corresponding bands in lanes 1 and 5, 5 and 9, and 9 and 14 would be joined. Once specimen band locations have been confirmed, geometric correction for every specimen band is performed by the system transparently to the user.
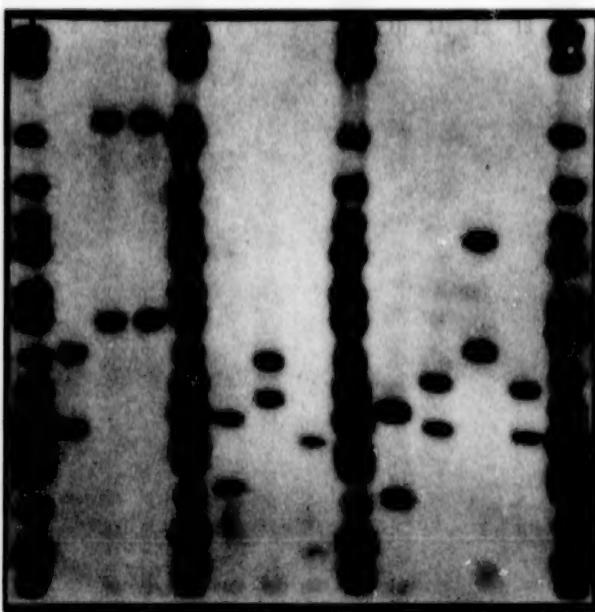
Figure 7. All bands in the four lanes containing size standards properly marked.


Figure 8. Band locations used to calculate specimen fragment lengths, and locations of the molecular weight size markers used for geometric correction and determination of fragment lengths.

Location of Specimen Bands

Location of bands in each of the specimen lanes is similar to the standards lanes, except that the operator is asked to enter the number of bands that are observed (Figure 8). If band relocation is required it is done just as it was in the lanes containing standards. Other identifying information, including tissue type, race of donor (if known), and case particulars, is entered at this time.

Calculation of Fragment Lengths

Fragment lengths are calculated using the virtual position of the known size markers which were determined earlier by geometric correction. Several models relating fragment length to electrophoretic mobility have been proposed in the literature (Elder and Southern 1983). Because the HaeIII enzyme system produces relatively short fragments, and a closely-spaced standard ladder is used, the simple logarithmic model is adequate. The calculated molecular weight values for each lane are written to disk for storage, data manipulation and possible incorporation into a data base. They are also included in a printed report for each autoradiogram (Table 1). A video hard copy is a useful addition to the case file because it shows the band locations which were used to calculate fragment lengths. Along with the numerical results, it can facilitate peer review and quality control of autoradiogram interpretation.

ACCURACY AND PRECISION OF DNA FRAGMENT LENGTH MEASUREMENT

Both accuracy and precision of fragment length measurement are strongly dependent on the size markers which are included on every analytical gel. Accuracy refers to the proximity of the average value of a measurement to the true value, while precision is the spread in individual measurements. Accuracy is composed of two components, precision and bias (systematic error). Precision is characterized by a symmetrical distribution about the mean, while bias is one-sided.

For an application where each value is measured only once or twice, it is essential that measurement precision be well characterized. This is particularly so for the case of forensic DNA fragment length measurement, where the limited quantity of DNA available precludes multiple analyses. Only when precision is known can it be determined whether or not two bands are indistinguishable within the limitations of the technique. Good precision allows the experimenter to determine the extent of systematic error (bias). For side-by-side comparisons, high accuracy is not strictly necessary; it is only required that the bias be repeatable. For comparisons over time, or for comparison with results obtained with other image analysis systems, however, it is desirable to eliminate measurement bias.

Table 1. COMPUTER PRINTOUT OF SPECI-MEN FRAGMENT LENGTHS MEASURED FROM THE BAND LOCATIONS SHOWN IN FIGURE 8.

CALCULATED FRAGMENT LENGTHS (log model)

Autoradiogram: TEST—DNA Probe: D2S44—MW Standard: Collab. Res.	Serologist: KLM:—IM. Analysis: 27–JUL–89— No. markers: 16
Lane 2: Blood /Stain, Black	,
Band 1 MW = 1954 bp	
Band 2 MW = 1406 bp	
Lane 3: Blood /Stain, Black	,
Band 1 MW = 6089 bp	
Band 2 MW = 2222 bp	
Lane 4: Blood /Stain, Black	, Duplicate of lane 3
Band 1 MW = 6071 bp	
Band 2 MW = 2212 bp	
Lane 6: Blood /Stain, Black	,
Band 1 MW = 1432 bp	
Band 2 MW = 1018 bp	
Lane 7: Blood /Stain, Black	,
Band 1 MW = 1880 bp	
Band 2 MW = 1575 bp	
Lane 8: Blood /Stain, Black	,
Band 1 MW = 1318 bp	
Band 2 MW = 745 bp	
Lane 10: Blood /Stain, Black	,
Band 1 MW = 1536 bp	
Band 2 MW = 1021 bp	
Lane 11: Blood /Stain, Black	,
Band 1 MW = 1783 bp	
Band 2 MW = 1456 bp	
Lane 12: Blood /Stain, Black	,
Band 1 MW = 3248 bp	
Band 2 MW = 2062 bp	
Lane 13: Blood /Stain, Black	, Duplicate of lane 11
Band 1 MW = 1784 bp	
Band 2 MW = 1446 bp	

System accuracy can be assessed over the entire molecular weight range of interest by measuring reference standards in exactly the same way as specimens are measured. Using one ladder as a standard, a different ladder, which was considered as an unknown, was measured. With the 31-band Lifecodes (Lifecodes Corp., Valhalla, NY) ladder as standard and the 19-band Collaborative Research (Collaborative Research Inc., Bedford, MA) ladder as unknown, all values were within 1% of the known value. When the Collaborative Research ladder was the standard, all values measured for the Lifecodes ladder were within 2.5% of the known value. The higher accuracy obtained using the Lifecodes ladder, which has more closely-spaced bands, was expected; any of the models which relate fragment length to observed electro-phoretic mobility perform best over narrow molecular weight ranges (Elder and Southern 1983).

The contribution of measurement imprecision to overall accuracy is itself composed of two com-ponents, attributable to the measurement system and to the specimen. The system imprecision was evaluated by comparing repeated measurements of the same autoradiogram, carried out by different operators over a period of several weeks and on separate work stations. The range of system imprecision was 0.1–0.3%.

Precision of fragment length determination was investigated for both ideal specimens and for case work specimens. The range of values in the accuracy study already discussed, using size markers as unknowns, was less than 1%. Several hundred human cell line standards, which are included on every case work autoradiogram as a quality control, varied by less than 2.5% from one another. We also found a variation of less than 2.5% between the fragment lengths measured on stains produced from blood exemplars from 100 victims of sexual assault and values measured in specimens deriving from vaginal swabs from the same victims. The latter were prepared by a differential extraction procedure which takes the advantage of the differences in membrane permeability between epithelial (vaginal) cells and spermatozoa (Gill *et al.* 1985).

Whichever size marker has been chosen, accuracy of fragment length measurement also depends on its proper use. It has already been mentioned that closely-spaced bands allow more accurate modeling of fragment length from electrophoretic mobility and that the ladder should span the range of expected fragment lengths. Furthermore, a size marker lane must be run every few lanes, since a linear geometric correction is employed. The FBI flanks every three or four specimen lanes with size markers. Finally, band intensity in marker lanes should be comparable to that in the specimen lanes so that precision of band location will be similar.

CONCLUSION

The VNTR technology, including the use of a semi-automated system to estimate fragment lengths for comparisons and calculation of the probability of occurrence of a given profile, provides the best method now available to exclude a falsely accused individual. Using the image analysis system we have described, it takes approximately four minutes to perform a complete analysis of a typical autoradiogram containing four standard lanes and ten specimen lanes. All of the operations are menu-driven, and an inexperienced user can generally begin to analyze data after a single demonstration.

Samples which have been collected from generally recognized racial subgroups have been used to generate a population data base for each locus analyzed. Measured fragment length values for these samples are transferred directly to the appropriate data base. The statistical methods established for calculating allele frequencies from these data are discussed in some detail elsewhere in these proceedings (Budowle and Monson 1990).

Based on our studies of measurement precision in ideal and case work specimens, we have adopted the criterion that two bands, which by preliminary visual comparison appear to match, are indistinguishable (by present techniques) if their fragment lengths differ by 2.5% or less. Of course, the examiner may still judge the comparison inconclusive, even when the numerical results fall within this narrow boundary. When a fixed bin approach is used to define alleles, the precision associated with each fragment length measurement in a specimen is also used to determine to which bin a band should be assigned for the purpose of calculating phenotype frequencies (Budowle and Monson 1990).

ACKNOWLEDGEMENT

REFERENCES

Budowle, B. (1990). A Protocol for RFLP Analysis of Forensic Biospecimens. In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C., pp. 57–62.

Budowle, B. and Monson, K. L. (1990). A Statistical Approach for VNTR Analysis. In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C., pp. 121–126.

Elder, J. K. and Southern, E. M. (1983). Measurement of DNA length by gel electrophoresis. II. Comparison of methods for relating mobility to fragment length, Anal. Biochem. 128:227–231.

Gill, P., Jeffries, A. J. and Werrett, D. J. (1985). Forensic applications of DNA fingerprints, Nature 318:577–579.

Monson, K. L. (1988). Semiautomated analysis of DNA autoradiograms, Crime Laboratory Digest 15:104–105.

Pun, T., Trus, B., Grossman, N., Lieve, L. and Eden, M. (1985). Computer Automated lanes detection and profiles evaluation of one-dimensional gel electrophoretic autoradiograms, Electrophoresis 6:268–274.

THE NATIONAL CRIME INFORMATION CENTER: PAST, PRESENT AND FUTURE

Donald M. Johnson

FBI Technical Services
Washington, D.C.

THE BEGINNING

In the mid-1960s, the law enforcement community throughout the United States was clamoring for a means to instantly access nationwide criminal justice information. The criminals of this era were more mobile than ever. Oftentimes a bank robber or murderer could be hundreds, even thousands, of miles from the scene of the crime within hours. The police officer of the 1960s was also witnessing an upsurge in the movement of stolen property from one jurisdiction to another. The need for better communications among all of law enforcement, city to city, state to state, and coast to coast, was evident. Law enforcement responded by establishing the National Crime Information Center or, as it is known to over a half million sworn and civilian criminal justice employees throughout the United States—NCIC.

As a result of a joint effort, spearheaded by the FBI and the International Association of Chiefs of Police (IACP), local, state, and Federal agencies began a nationwide criminal justice information system—a system that allowed law enforcement to share complete and accurate information with other members of the criminal justice community in an on-line, real-time fashion. When NCIC's power switch was turned on in January, 1967, over 22 years ago, the greatest single example of local, state, and Federal criminal justice cooperation, the NCIC embarked on its mission—to assist the United States law enforcement community with documented criminal justice information regarding stolen vehicles, stolen license plates, stolen articles, stolen guns and wanted persons.

THE TECHNOLOGY

The NCIC is a national criminal justice information system. It is supported by computer equipment located at FBI Headquarters in Washington, D.C. A National Advance Systems, Model 9080 multiprocessor with 64 megabytes of main memory host the system. Telecommunications are handled by Computer Communication Inc., Model 85, front end processors. This supports a telecommunication network of 138 lines, a majority of which are 4800

and 2400 baud bisynchronous contention circuits. All lines are dedicated, and the long-line circuits are multiplexed into 9600 baud trunk lines. Tall circuits and short-line circuits are point to point, and a few circuits are less than 150 baud with either single or multiple terminals on each. The NCIC host telecommunications network of 138 lines supports 64 locations in the 50 states, District of Columbia, Commonwealth of Puerto Rico, United States Virgin Islands and Canada. Once supplied with the main line from NCIC, each state operates an independent system that interfaces with NCIC through state computer interface agencies known as Control Terminal Agency (CTA). The CTA is usually the State Police or State Department of Public Safety. In turn, the CTAs operate intrastate criminal justice networks, which are largely responsible for making NCIC criminal justice information available to law enforcement agencies throughout the country.

THE MANAGEMENT

While the day-to-day management of NCIC rests with the FBI, management responsibilities are shared with an Advisory Policy Board (APB) which makes recommendations to the Director of the FBI regarding philosophy, concept and policy relevant to the day-to-day operations of NCIC. The 30-member APB consists of 20 representatives elected by or from the states, six members appointed by the FBI Director and four representatives from National Criminal Justice Associations. The growing needs of law enforcement are of paramount importance to the APB and the FBI. Historically, the FBI Director has approved and adopted virtually all of the APB's recommendations.

Additionally, each of the 50 states and selected Federal agencies also share in the management of NCIC through the CTA. Through four Regional Working Groups (RWG), every state has a representative that meets to provide technical and policy recommendations for consideration by the APB. Each RWG also consists of six elected representatives from local criminal justice agencies. In order

to participate in NCIC, each CTA must comply with NCIC's standards for security, training, auditing, data quality and response time. While the direct line provided to each state is provided at no charge, the costs of operating NCIC from the state level on are borne by the state and/or the end users of NCIC.

THE GROWTH

The NCIC has grown from five files in 1967 to 12 files in 1989. As the information needs of law enforcement have increased over the years, NCIC has responded by adding criminal justice information files for Missing Persons, Unidentified Persons, Stolen Securities, Secret Service Protective Information, Foreign Fugitives and the Interstate Identification Index (the computerized criminal history of over 12 million individuals, verified and authenticated by fingerprints). In 1968, NCIC handled over six million total transactions. In 1988 alone, criminal justice agencies queried NCIC over 228 million times. The national files have grown from several hundred thousand available records in 1968 to over 20 million records in 1989, instantly available to law enforcement. Even with such tremendous growth, NCIC has maintained its self-imposed standards of responding to hot file transactions within two seconds and five seconds for criminal history inquiries—a phenomenal accomplishment for a national information system the size of NCIC.

THE FLEXIBILITY

The responsiveness of NCIC to the changing needs of law enforcement can be exemplified by NCIC's technological and philosophical flexibility to introduce new and innovative law enforcement files. In 1975, the number of runaway teenagers, of abused children, and of kidnap victims was increasing at an alarming rate. NCIC created the Missing Persons File allowing criminal justice agencies to enter juvenile and adult missing persons. Since the inception of the Missing Persons File, the number of missing persons entered and removed from NCIC has grown over 300 percent.

The United States Secret Service (USSS) Protective File was activated in 1983 to assist the USSS with the demanding responsibility of protecting the President, Vice President and other United States officials. The file was established specifically at the request of the USSS to help determine the whereabouts of those individuals considered a serious threat to the President of the United States or any of their other protectees. An inquiry on a person entered in this file is instantly referred to USSS Headquarters in Washington, D. C., and the local agency is immediately contacted to determine the status of the person identified in the inquiry. The USSS Protective File is NCIC's smallest file containing an average of 30 individuals in it at any given time, but it has been extremely effective. During the six years this file has been in operation, NCIC users have located 572 individuals considered to be a threat to the President of the United States. The accomplishments of this data file alone attest to the outstanding reputation that NCIC has for adaptability, flexibility and effectiveness.

The ability of NCIC to adjust to the ever-changing and ever-increasing needs of law enforcement are exceptional. With the need for the addition of fingerprint classification information to be added to NCIC's wanted person records, the FBI's Identification Division and NCIC joined hands to convert the Henry Classification to an alphanumeric which could be entered and searched in NCIC. Over the years, NCIC has added dental characteristics, blood types and scars/marks/tattoos, further demonstrating the flexibility of NCIC. The conversion of standardized DNA information to a searchable alphanumeric is a positive extension of the capabilities of NCIC.

THE FUTURE

Since 1986, a federally funded study of NCIC was conducted to develop and implement a new NCIC which will meet enhanced user community requirements through the year 2000. This study, called NCIC 2000, has resulted in the adoption of over 75 concepts by the APB and the Director of the FBI. The implementation process of NCIC 2000 now begins. The increased functionality resulting from the NCIC 2000 study consists of improved data quality, positive identification through the transmission of fingerprint and photographic images, usage statistics, intrusion detection, enhanced flexibility in the searching of the NCIC data bases, the development of work stations, enhanced hit responses, on-line validations, artificial intelligence and many other critical improvements to today's NCIC. With growing management, administration, and congressional support, the FBI now moves the NCIC 2000 project into the funding process.

With the implementation of NCIC 2000, the NCIC looks forward to advanced and expanded functionality, as well as improved data quality and system integrity. The transmission and receipt of photographic and fingerprint images vastly im-

proves the positive identification capabilities of NCIC.

The NCIC stands on the threshold of revolutionizing the national information system capabilities of America's law enforcement community. The addition of a National index for DNA characteristics, or the addition of DNA fields to existing NCIC records could provide law enforcement with on-line entry and searching capabilities of DNA information. The magnitude of DNA as a powerful investigative resource provided by the forensic community is recognized by NCIC. The FBI, the NCIC Staff and the NCIC APB stand ready to support whatever developmental resources are needed to assist the scientific community in bringing this valuable investigative tool to the aid of law enforcement in the on-line, real-time world of NCIC.

BLANK PAGE

DNA TYPING IN THE ROYAL CANADIAN MOUNTED POLICE

Ron M. Fourney, Gary G. Shutler, Nancy Monteith, Lucy Bishop, Barry Gaudette and John S. Waye

Central Forensic Laboratory
Royal Canadian Mounted Police
Ottawa, Ontario, Canada

In 1980, Wyman and White discovered the first highly polymorphic DNA marker in the human genome (variable number tandem repeats), and Botstein *et al.* (1980) reported that restriction fragment length polymorphisms (RFLPs) could be used to construct a genetic linkage map. These events set the foundation to utilize DNA polymorphisms to investigate genetic variation. The utility of this technology has been exploited for disease diagnosis (Lerman 1986) and more recently attracts the forensic investigator with the ultimate power of discrimination, the ability to identify a single person to the effective exclusion of all other individuals.

On April 5, 1989, genetic evidence in the form of DNA typing was entered into the criminal courts in Ottawa and in so doing marked the precedent setting case for the Royal Canadian Mounted Police (RCMP) forensic laboratory (Regina vs McNally). The introduction of this technology to forensic science has involved many considerations and still remains on a flexible schedule of adaptation that is attuned to the rapid advances in the field of molecular genetics and recombinant DNA technology. Many changes that will occur in the future will involve the training and proficiency testing of forensic scientists so that DNA typing can eventually be implemented into our seven regional service laboratories throughout Canada. Currently, the RCMP forensic laboratories offer assistance to all police organizations in the country. The RCMP has a special role in police matters involving Federal jurisdiction and contracts police service in eight out of 10 provinces. The following report details some of the initial considerations dealt with prior to the introduction of this technology into our laboratories. In particular, multilocus versus single locus analysis, restriction enzyme and DNA probe selection are discussed in conjunction with the practical aspects and the future prospects of the technology.

Multilocus versus Single Locus Analysis

Variable number of tandem repeats and minisatellites, hereafter referred to as VNTRs, are short arrays of core repeat oligonucleotide sequences organized as several thousand interspersed loci. Depending on hybridization conditions, VNTR probes may recognize either a single locus (Wyman and White 1980; Nakamura *et al.* 1987a) or simultaneously detect allelic variation at multiple loci on several different chromosomes (Jeffreys *et al.* 1985ab, 1986). In the absence of blocking DNA (that is, salmon or herring sperm DNA), low stringency hybridization adjusted through ionic strength, formamide concentration or the use of RNA probes (Fowler *et al.* 1988) in hybridization mixture and/or temperature of hybridization will allow probes to recognize and bind to loci sharing molecular similarities that have 30–40% mismatch between the probe and target sequence. As shown in Figure 1A, the DNA probe YNH24 (D2S44, Nakamura *et al.* 1987b) recognizes concurrently a great diversity of bands in a Southern transfer when hybridized under relaxed or less stringent conditions. However, the same membrane hybridized under more stringent conditions with the D2S44 probe, generates a locus-specific allelic pattern (Figure 1B). The best known multilocus probes, namely 33.15 and 33.6 identified by Alec Jeffreys and co-workers (Weller *et al.* 1984; Jeffreys *et al.* 1985ab, 1986) were originally derived using the repeat sequences associated with the myoglobin gene locus. One of these probes has been reported (Gill *et al.* 1985) to provide individual discrimination power ranging from 0.2 to 2.4 x 10^{11} thereby providing the forensic scientist with a powerful tool for establishing identity accomplished through the process of a single hybridization.

VNTR probes used under highly stringent hybridization conditions recognize a single locus. By consequence, genomic DNA digested with appropriate restriction enzymes that flank VNTR loci and probed under locus-specific conditions generate a maximum of two allelic band fragments for each locus. The heterozygous pattern of inheritance will present as two bands and the homozygous individual will show a single intense band. The potential number of target loci is greatly reduced using this approach but individualization is

YNH24
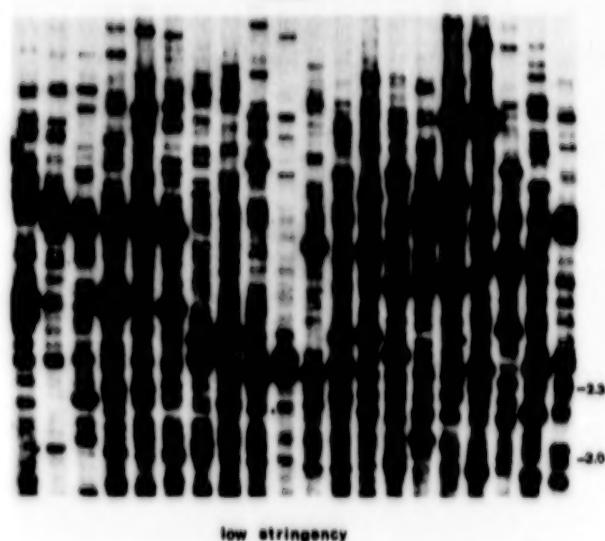
low stringency

Figure 1A. Low stringency hybridization was carried out on genomic DNA isolated from twenty individuals. After *Hae*III digestion, samples were electrophoresed in 1X TBE, 1.0 % agarose gels, transferred to a nylon membrane and probed with P[32] labelled pYNH24 (D2S44 locus) at 52°C in a hybridization solution consisting of 25% formamide, 6X SSC, 10% dextran sulfate, and 10 X Denhardt's solution. The membrane was washed at high stringency (0.1X SSC, and 0.1% SDS at 65°C) and autoradiographed at -70°C for 24 hr using Dupont Cronex Lightning Plus Screens and Kodak X-AR5 x-ray film.

easily achieved through multiple probings of the same membrane with several distinct probes recognizing well defined and genetically independent loci.

Although the principal interest of DNA typing to forensic scientists is its power of individual discrimination, the final results must be defendable in court. The net goal of all forensic experts is to present results in a clear and concise manner such that the interpretation is understood and accepted with confidence. Ideally, the DNA evidence introduced into court should be easily interpreted and comparison of RFLP patterns should be readily visible as a match or mismatch to the lay audience comprising the jury. As noted previously, multilocus probes generate a complicated autoradiographic picture often comprised of 10–60 bands of varying intensities. Differences in band intensity is due to varying amounts of homology between the probe and heterologous target loci. Single locus probes generate a one or two band allelic pattern on autoradiograms that is readily interpreted by classical Mendelian inheritance and genetic principles and is amenable to standard statistical analysis. Collectively, these attributes should allow the expert witness to make an effective presentation of the results.

Enhanced sensitivity and specificity are major advantages of single locus probe analysis. Single locus probes require sub microgram quantities of genomic DNA for genetic analysis (Figure 1B). In our experience as well as that of others (Budowle *et al*. 1988; Werrett 1987), multilocus probes require larger amounts of genomic DNA, typically in the 1–2 μg range, a potentially limiting factor in many forensic samples. Moreover, intact high molecular weight DNA is required for optimal multilocus detection while single locus probes can be targeted specifically to DNA fragments in a lower molecular weight range, thus accommodating DNA samples which are partially degraded.

YNH24

high stringency

Figure 1B. High stringency hybridization was carried out using the same membrane noted in Figure 1A. All conditions were the same except the hybridization buffer contained 50% formamide and 3X SSC instead of 25% and 6X SSC.

The ability of multilocus probes to recognize genomic DNA from both closely and distantly related species has proven useful in establishing identity (Burke and Bruford 1987; Hill 1987; Jeffreys and Morton 1987; Welton *et al*. 1987; George *et al*. 1988; Ryskov *et al*. 1988). However this characteristic may limit their usefulness in forensic analysis. For example, biological samples collected as evidence are often contaminated with bacteria and may be composed of stains from several species. Consequently, probes detecting DNA sequences in different species could make interpretation of results very difficult. The ability to control the specificity of probe recognition is a highly desirable feature for DNA typing.

From a forensic point of view, the routine investigation of mixed evidential samples has often been the supreme challenge of each practicing se-

rologist. Mixed samples such as blood stains contributed from two individuals or sperm and vaginal cell mixtures lead to complex multiband patterns of 80–200 bands using multilocus probes. In contrast, single locus probe autoradiographic patterns are composed of at most a four band pattern attributable to two individuals. Although methods exist to separate spermatozoa DNA from vaginal and other cells no methods exist for separation of mixed blood stains. A discrete number of bands generated from a mixed blood stain analysis is viewed as an easier pattern to comprehend and explain.

Current investigations suggest that DNA typing may become highly applicable to partially degraded DNA (unpublished data). These efforts could potentially unlock important new evidence in old cases as well as provide valuable information from recent casework where evidential samples have been exposed to natural degradation. Degraded DNA restricted with standard enzymes results in the production of low molecular weight fragments (1–4 kilobase (Kb)) that lack many of the more informative alleles (4–20 Kb fragments) detected by multilocus analysis. In addition the low molecular weight fragments detected by multilocus analysis form a compact and intense autoradiographic pattern which is difficult to interpret. Fortunately, many single locus probes generate allelic patterns in the range of 0.5–4 Kb and may provide a useful means for examining partially degraded material. The information content of these samples can be maximized by running high concentration agarose gels with special refined agarose (NuSieve GTG Agarose, FMC) that enhances resolution in the low molecular weight range. Through the judicious use of single locus probes that recognize a low molecular weight series of alleles, DNA typing should be successful for many degraded specimens.

A future advantage for using single locus probes will not become readily apparent until autoscan systems and computer databases are in common use. Automated scanning of multiband patterns using electronic imaging equipment has been hampered due to limitations in electrophoretic resolution as well as resolution limits in camera optics for digitization of electronic signals. One or two band patterns per sample lane are easily detected and managed by electronic imaging equipment and therefore provides the researcher with an important tool to rapidly and accurately assess VNTR population data base information on specific loci in a highly reproducible manner.

RESTRICTION ENZYME SELECTION, DNA PROBES AND STANDARDIZATION

Restriction Enzymes

The RFLPs revealed by VNTR probes rely on the detection of size variation in the tandem sequence arrays confined within fixed restriction sites that flank the target area. As a consequence, polymorphic VNTRs can be generated by using a number of different restriction enzymes provided that there is no cleavage site within the VNTR array. In collaboration with the Federal Bureau of Investigation (FBI), the RCMP has carefully selected the following characteristics as important requirements for a restriction enzyme useful for forensic DNA typing: (1) must cleave in the flanking sequence of several genetically independent and highly polymorphic loci, (2) the restricted fragments must be easily resolved by conventional agarose gel electrophoresis, (3) the enzyme must effectively digest DNA under a wide variety of conditions, (4) common mammalian methylation patterns should not affect enzyme restriction and (5) the enzyme should be inexpensive. All points have been discussed in detail elsewhere (Budowle *et al*. 1990) but the following represents a short synopsis of our criteria.

Although several enzymes have been used in the past to create allele frequency population DNA data bases (*Pst*I, *Hinf*I, *Pvu*II and *Alu*I), *Hae*III has been chosen as the enzyme of choice for our forensic program. *Hae*III is a frequent cutter that recognizes the four base sequence GGCC and produces many DNA fragments in a size range directly amenable to DNA typing in our system. For example, D2S44 (probe pYNH24, Nakamura *et al*. 1987b) is a highly polymorphic VNTR locus which has become one of the core loci in many forensic programs (heterozygosity index of 0.97). In preliminary data base analysis of approximately 700 nonrelated Caucasians (sampled from the vicinity of Ottawa) it became evident that the enzyme *Hae*III produces its vast majority of allele fragments in the 1–4 Kb size range (Figure 1B). Consequently this results in the production of DNA fragments in a size range easily resolved with conventional 1% agarose gel electrophoresis in a 20 x 14 cm (w/l) gel format. In a collaborative study initiated in the forensic laboratories of the FBI (Budowle *et al*. 1990) it became apparent that other enzymes such as *Alu*I, *Hinf*I and *Pst*I did not create as many D2S44 homologous fragments in the 1–4 Kb range as *Hae*III.

Forensic applications of DNA require a robust restriction enzyme. Samples entering the forensic laboratory are far from being pristine biological specimens and in some cases are contaminated with many known and unknown foreign substances which could potentially co-purify with the DNA. In general there are few substances which alter the integrity of DNA but there are many inhibitors of enzymes which could affect DNA digestion and create problems in interpretation of the final hybridization results. Consequently the selection of a robust restriction enzyme that is capable of working efficiently under suboptimal conditions is regarded as a definite advantage. As documented in the literature (Blakesley *et al.* 1977) and in experiments conducted in both the FBI and RCMP forensic laboratories (Budowle *et al.* 1990) *Hae*III has demonstrated complete and efficient digestions in a broad temperature range (18°C to 70°C), in different ionic strengths (25 mM to 150 mM) and a range of pH (6–9). Furthermore the forensic applicability of this enzyme has been recently tested. Adams (1988) extracted DNA from many biological specimens previously subjected to contamination from a wide variety of domestic substances such as unleaded gasoline, motor oil, detergents, acids, bases, salts and bleach. In all cases DNA was successfully extracted and restricted efficiently with *Hae*III. Although this study was designed to test the feasibility of working with biological specimens that have been exposed to common chemical products it does suggest that contaminating trace elements that often co-isolate during DNA purification, do not affect the activity of this enzyme. Even though extraction procedures are designed to yield highly purified DNA, it has been noted that traces of organic solvents left in crude preparations of DNA do not appreciably alter *Hae*III activity (unpublished data). *Hae*III appears to tolerate high concentrations of genomic DNA and is unaffected by elevated protein concentrations such as BSA or commonly employed extraction buffer ingredients such as urea. All of the former ingredients have been reported to affect the activity of many commonly employed restriction enzymes and subsequently may lead to incomplete DNA digestion patterns.

Many restriction enzymes are sensitive to the methylation of their respective sequence recognition sites. The potential result of this condition could be the inhibition or reduction of restriction enzyme activity. If the methylation is tissue specific, this could subsequently produce RFLP variation in DNA isolated from different tissues from the same individual (Adams and Burdon 1985; Silva and White 1988). Since the forensic scientist will be called upon to compare DNA profiles from a variety of tissues, it is mandatory that restriction enzymes utilized in DNA typing be unaffected by natural biological variation such as methylation. In higher eukaryotes such as mammals, methylation of the 5′ cytosine residue that immediately precedes a guanine residue is the most frequent modification identified. Fortunately this sensitive CpG dinucleotide is not present in the recognition target sequence of *Hae*III (5′-GGCC-3′) and when the *Hae*III site is followed by a G residue (GGCCG) methylation of the 3′ C does not affect *Hae*III activity. Therefore methylation of *Hae*III digested DNA would not expect to play a role in causing genetic variation. In contrast other endonucleases have demonstrated reduced activity due to methylation (Huang *et al.* 1982; McClelland and Nelson 1988) which could ultimately become a source of error.

An enzyme used in a DNA typing program should be cost-effective with respect to the initial monetary cost and the concentration required to achieve a complete digestion. Although not a major concern (considering the amount of enzyme used per analysis), the selection of *Hae*III has the added bonus of being a very inexpensive restriction endonuclease. As a final note, this enzyme's excellent stability and activity under suboptimal conditions would suggest that it is a cost effective restriction enzyme for a forensic DNA typing program.

Probe Selection

What makes a good forensic DNA probe? In addition to recognizing polymorphisms, forensic probes should identify loci that satisfy the following criteria: (1) DNA loci must have documented mutation rates, (2) RFLPs detected by the probes must exhibit Mendelian inheritance patterns, (3) chromosome location should be recorded by either the Yale Gene Library or the International Human Gene Mapping Workshops, (4) polymorphic loci should be documented in the scientific literature and the nature of polymorphism identified (that is, restriction site polymorphisms, insertion and deletion, VNTR), (5) all probes should be available for confirmatory testing by several independent laboratories and (6) a population data base of allelic frequencies must be made available to all laboratories for comparative purposes.

Several other considerations should be noted. Probes should be free of repetitive sequence re-

gions which would elevate the hybridization background significantly. The ideal forensic probe should be carefully cloned to provide a DNA sequence that is human specific (or higher primate) and completely free of cross-hybridization with DNA from bacteria or other species.

All probes being examined in the forensic program in Ottawa conform to the above criteria and are compatible with the endonuclease *Hae*III and our electrophoresis system. The primary probes in current use (D2S44, D1S7, D16S85, D4S139 and D17S79) comprise our regional data base. These probes fall within the resolution range of our electrophoresis system and are amenable to automated digitized band sorting and entry into a computer accessible data base. It is fully expected that with development, several new probes that detect bands in the 0.8–2.3 Kb range (that is, D11S129, D20S15) could become important tools in the analysis of DNA that is partially degraded. As more VNTR loci are discovered it is important that all potential forensic probes be carefully evaluated and that a forensic DNA typing program be flexible to accommodate additions and changes.

There are many practical considerations concerning probe selection. A prime concern of the RCMP is to establish a DNA typing program that is compatible with other North American law enforcement agencies. One major advantage of adopting a standardized approach to probes, enzymes and statistical analysis is that the DNA typing process will be constantly peer-reviewed and receive court precedence in several centers. It is crucial for the successful establishment of DNA typing as a legitimate forensic procedure, that the scientific and forensic communities have equal access to probes. This will encourage careful evaluation of DNA probe characteristics, validation of population data bases and publication of data in the scientific literature. Consequently, probes must be readily available and economically priced to attract the vast majority of the forensic and clinical genetics community.

Practical Application of DNA Typing at the RCMP

Casework conducted by the RCMP forensic laboratories represents a team effort. Prior to the entry of evidential samples into the molecular genetics laboratory a number of specialists representing such diverse disciplines as hair and fiber, toxicology and serology have on most occasions initiated the scientific evaluation of the collective evidence. Generally, biological samples destined for DNA analysis will be prescreened by serology spe-

cialists. As a precautionary measure to ensure extraction of high molecular weight DNA, the principal crime scene investigator and all subsequent participating specialists are encouraged to freeze biological specimens destined for DNA typing.

Historically DNA typing in our own laboratory as well as other forensic centers, represents pivotal associative evidence in very serious crimes involving sexual assault and/or homicide. In order to make DNA typing more meaningful, it is crucial that known relevant biological samples from suspects, victims, and others associated with the crime be available for complete analysis. Unlike American jurisdictions, there is no statutory authority in Canadian law for obtaining biological control samples such as hair or blood from any individuals accused of a criminal act. The sole exception is a provision in the law to obtain blood samples from persons accused of impaired driving and these samples can only be used to determine blood alcohol content. Consequently, the RCMP will endeavour to examine and establish important DNA typing information to link serial crimes or to establish the identity of unidentified human remains, and will process criminal cases where known samples of biological materials are obtained voluntarily from victims and/or suspects.

The format of DNA typing at the RCMP laboratories parallels that of many forensic centers. Typically biological stains (that is, minimal volumes; 5–50 μl of blood, 1–10 μl of semen) and control blood samples (100–700 μl) are processed to obtain purified DNA. Although numerous DNA isolation procedures have been described (Gross-Bellard *et al*. 1973; Cory *et al*. 1983; Maniatis *et al*. 1982; Law *et al*. 1984; Weeks *et al*. 1986; Bowtell 1987; Longmire *et al*. 1987; Reymond 1987) it has been our experience that purified high molecular weight DNA, suitable for forensic analysis is most efficiently obtained by enzymatic Proteinase K treatment followed by organic extractions. Our current protocol (modified from Fourney *et al*. 1989) has been designed to handle multiple samples containing minimal amounts of starting material and to require only basic equipment found in most clinical or forensic laboratories. Following the isolation of DNA, the concentration is determined by hybridization with a probe that recognizes highly repetitive human sequences (detailed in the following discussion) and providing that the sample is not limiting, the integrity is documented by mini-gel electrophoresis. Since resolution of autoradiographic bands is largely dependent on the quantity of DNA analyzed, it is important to maintain a DNA

sample load below 5 µg. In our experience the most accurate assessment of human DNA concentration has been carried by the direct quantitation of hybridization signal from serially diluted questioned samples and corresponding controls. Essentially DNA is serially diluted, slot blot fixed in alkaline buffer to membranes and probed with a highly sensitive and primate specific alpha satellite DNA probe (p17H8 detecting locus D17Z1; Waye *et al.* 1989; Figure 2). In contrast to fluorescent staining or spectrophotometric procedures which measure total nucleic acid content in a sample, this protocol is not sensitive to contaminating DNA from other species including bacterial and therefore will not produce a false over-estimation of human DNA concentration. The entire procedure can be completed in less than four hours (including autoradiographic exposure) and will determine only the amount of human DNA present in a mixed sample that is available for hybridization with forensic DNA probes.

Once samples have been digested with *Hae*III, the questioned samples and a control DNA flanked by lanes of molecular weight marker are subjected to electrophoresis in a 1% agarose gel prepared in 1X TBE. Gel size, sample format and markers all conform to a standard pattern that results in an autoradiogram that can easily be scanned and digitized with a software package developed by the FBI (Monson 1988). Following electrophoresis, the gels are briefly stained with ethidium bromide (10 min), photographed under 312 nm fluorescence and transferred to nylon membrane using conventional procedures (Southern 1975). Our current transfer procedure employs a neutral charged membrane which results in low backgrounds and excellent durability for successive reprobing. Recent studies conducted in our laboratory with conventional alkaline transfer procedures (Reed and Mann 1985; Fourney *et al.* 1989) suggest that some enhanced sensitivity can be achieved with specific forensic VNTR probes, and a change to alkaline transfer may be warranted.
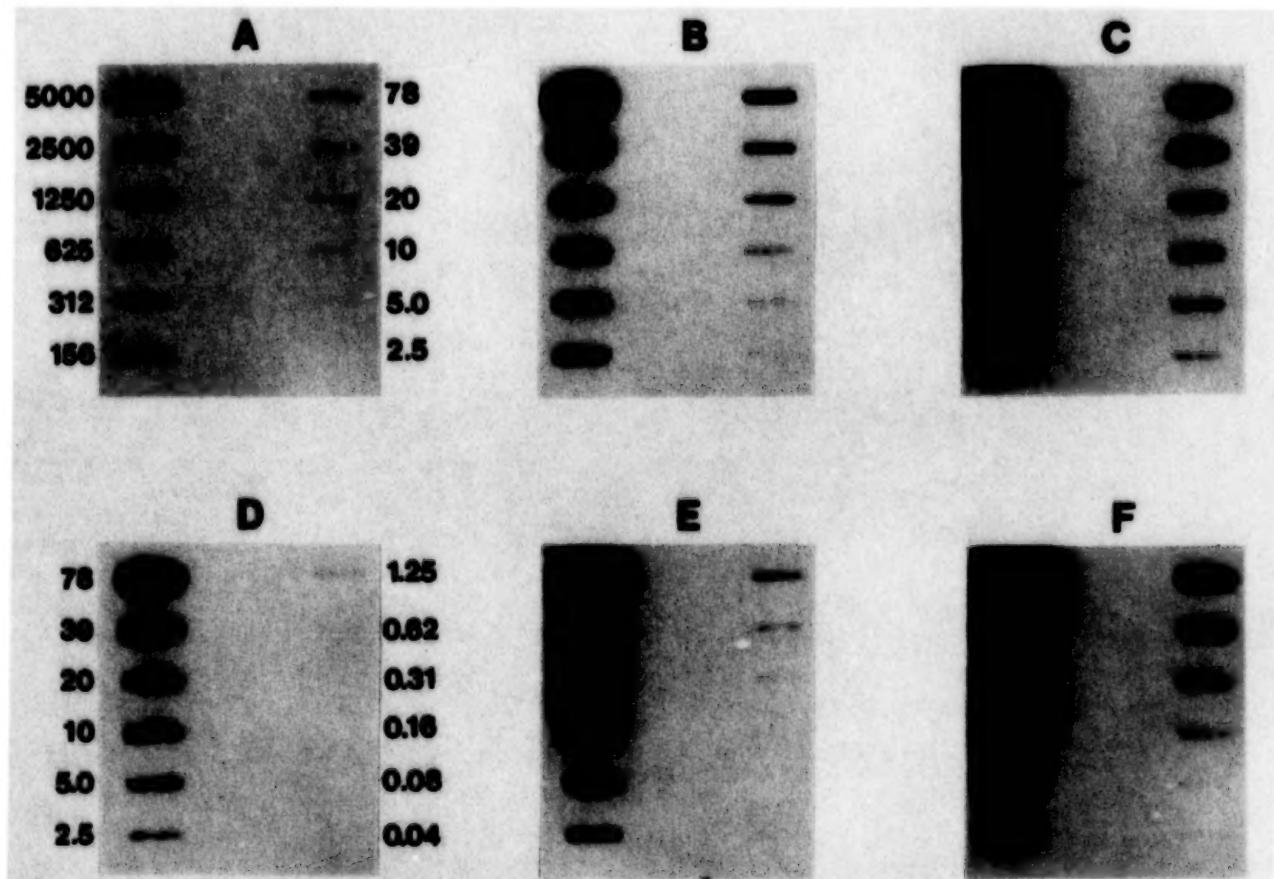

Figure 2. Sensitivity of slot blot quantitation of human genomic DNA using the higher primate specific probe p17H8. Serially diluted human DNA was applied by alkaline fixation to a nylon membrane (DNA concentration indicated in ng), hybridized with p17H8 and autoradiographed with x-ray X–AR5 films and Dupont Cronex Lightning Plus screens under the following conditions: Panel A room temperature for 5 min, and panels B,C,D,E, and F for 15,30,60, 90 min and 12 hr respectively at −70° C.

It should be noted that all procedures used in the RCMP laboratory exclude the intercalating dye ethidium bromide from the gel and running buffer during electrophoresis. All staining of DNA is done after electrophoresis has been completed. It has been our experience that agarose gel electrophoresis of linear duplex DNA in the presence of ethidium bromide has a marked effect on the mobility of genomic DNA fragments detected by Southern hybridization (Waye and Fourney 1989). This mobility variation could lead to problems in the comparison of RFLP patterns and in the establishment of accurate allele frequencies in population data bases. In addition safety and disposal concerns should warrant that ethidium bromide be carefully monitored and limited to small volumes of solution.

Judicious use of a series of four to five VNTR probes quickly builds a DNA profile that is highly individualistic. The last probings in our system are always D4S139, DYZ1 (a monomorphic Y specific locus) and D7Z2 (a monomorphic autosomal locus). These probes have a high sensitivity and are capable of counteracting any potential loss in DNA from the membrane that could incur during the stripping and probing processes. The probe pMGB7 detects a 2731 bp HaeIII repeat fragment on chromosome 7 (D7Z2, Waye et al. 1987) and serves as an internal control for evaluation of electrophoretic mobility and the accuracy of molecular size calculations. A special note in DNA typing procedures is to use a control DNA on each blot that has been derived from a male donor. This not only establishes a marker lane with a known polymorphic pattern for each probe but will act as a positive control for sex determination. Prehybridization, hybridization and autoradiography is carried out as detailed in the legend of Figure 1B.

Once typed, the statistical significance of the DNA profiles is assessed with a large allele frequency data base (that is, one data base in current use has been generated from ~700 Caucasians from the Ottawa area). For all casework and population data base samples, the electrophoretic migration of DNA samples relative to molecular weight markers is assessed by comparison of the 2731 bp HaeIII fragment detected by the probe D7Z2. Evaluation of an internal monomorphic band of known size enables the investigator to detect abnormal mobilities in each of the DNA samples. The data base has been formulated using a conservative binning procedure established in collaboration with the FBI research laboratories. To date the major criticism of this method has been the conservative nature of the final results. However, a conservative presentation in data management of DNA typing results can be defended in court.

Future Perspectives

It is difficult to anticipate the future applications of molecular genetics in the field of forensic science since new discoveries in molecular genetics and recombinant DNA methodology appear on a daily basis. The current DNA typing program in the RCMP laboratories is due to expand rapidly over the next year and several major considerations will become paramount for successful establishment of a national police service.

First and foremost is the aspect of training and quality assurance. Since the introduction of DNA evidence will have a major impact on the outcome of many criminal trials in Canada, it is important that all procedures be carried out with a high degree of accuracy and reproducibility. It is the intention of the RCMP laboratories to introduce a rigorous training program for forensic DNA typing. Training will involve complete familiarization with molecular genetic principles and recombinant DNA technology as it applies to human genetics. New members to the forensic laboratory staff recruited for the program will have prior experience in molecular genetics and will be trained as forensic scientists. In addition, the service component of our program will be supplemented with experienced forensic scientists who will be trained in DNA technology. All training and development will be coupled with an active in-house research program.

Crucial to the success of DNA typing will be the acceptance of quality assurance guidelines, that address such aspects of DNA typing as controls, appropriate standards, quality of equipment and chemicals, treatment of data and the constant monitoring of accuracy and reproducibility as established through proficiency tests and audits. The ability to design a flexible set of guidelines which reflects the constant and rapid changes in molecular genetics is not a trivial task. Currently forensic centers throughout North America meet on a regular basis in a FBI sponsored Technical Working Group on DNA Analysis (TWGDAM). As a participant in this program, it is anticipated that the RCMP laboratories and the majority of forensic centers using DNA typing will adopt TWGDAM sponsored guidelines on data base management, quality assurance and training.

Future projects will also include evaluation of new genetic typing methods such as polymerase

chain reaction (PCR) (Saiki *et al.* 1985; 1988; Higuchi 1988). At present, this technology promises many advantages but it is fully recognized that limitations may exist due to contamination and/or Taq polymerase fidelity (White *et al.* 1989). PCR technology involving amplification of VNTR length polymorphism is a promising prospect for forensic analysis. It is anticipated that, a combination of both PCR length polymorphism and RFLP DNA typing will represent a powerful combined method for individual discrimination.

As an ongoing project, our laboratory will continue to expand the VNTR population data base. Currently we have short term plans to include an additional 500 individuals into the Caucasian data base. In addition, data bases for major ethnic and racial subgroups in the Canadian population as well as Native indigenous citizens are being developed. As noted in Figure 3, preliminary information involving the D2S44 probe, demonstrates a significant alteration in the allele frequency, between Native Canadians and Caucasians. Since the mandate of the RCMP includes a police service to a large number of indigenous citizens, preliminary studies will be expanded and a separate data base will be created for Native Indian DNA analysis. It is important to note that DNA typing of a Native Indian DNA sample using existing Caucasian data statistics could very well bias the final frequency calculation and constitute both a scientific and legal injustice (see Bin 8, Figure 3).

With the expansion of DNA data bases, it will become compulsory to enhance our computer access to this information. Expansion of data bases and links to a North American network will increase the capability of the investigator to conclusively determine the identity of human remains and perhaps generate important data in serial crimes. The success of this program will primarily rest on international cooperation and adoption of a standardized method of data assessment.
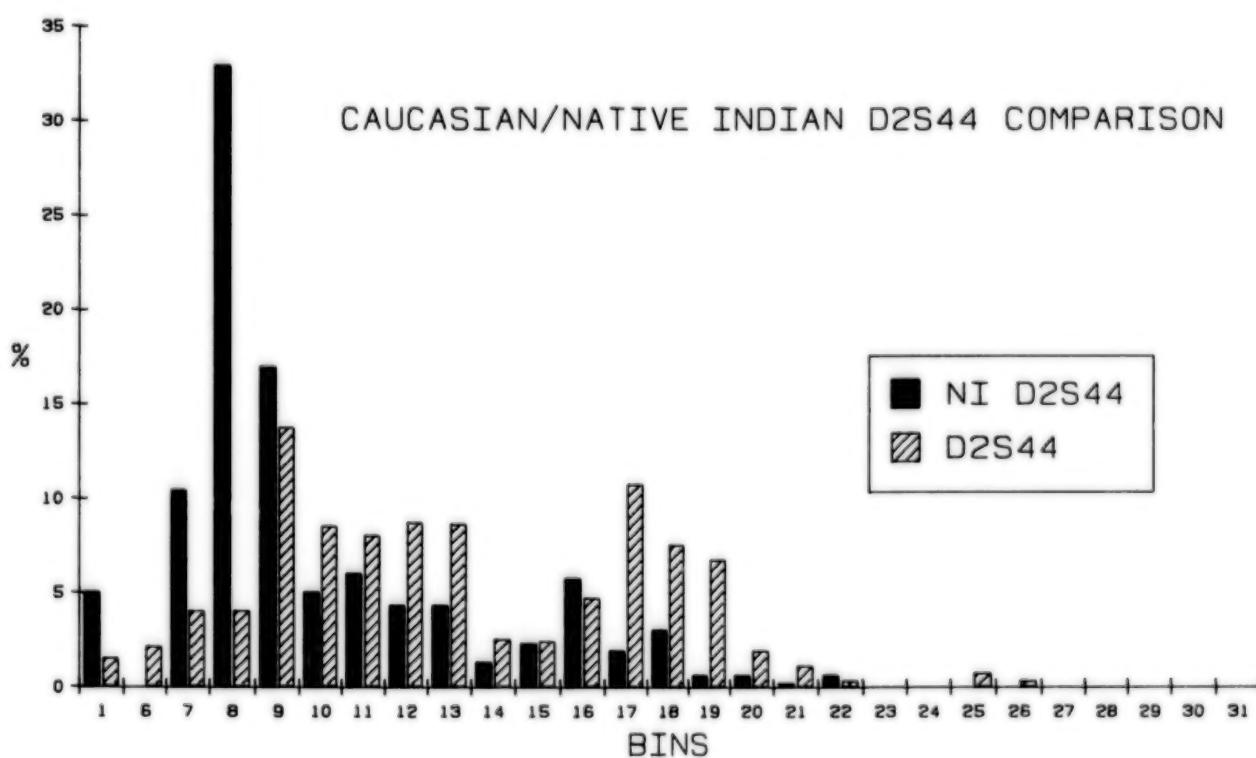

Figure 3. Allele frequency distribution of D2S44 locus in *Hae*III digested DNA from Caucasians and Canadian Native Indians. Caucasian data base has been established from 700 individuals and the Canadian Indian data compiled from 250 members. Size bins have been established using the FBI binning approach and the data compiled from automated digitized scanning of the autoradiograms using labeled markers of known size as reference points (Monson 1988). (Bin range: Bin 1-5, 0-1077, Bin 31>12829).

CONCLUSION

The purpose of this report has been to introduce the DNA typing service that is presently offered by the RCMP. As noted the methodology represents modification of procedures currently employed in clinical genetic laboratories throughout the world. Perhaps the major hurdle that DNA typing will face in Canada will be familiarizing the legal community and the general public to this new forensic technology. To date, DNA typing has largely remained within the realm of research laboratories and a select handful of private and forensic laboratories. The adoption of universal guidelines, a high degree of proficiency and quality control and an effective description of all procedures and principles should provide the momentum necessary for passing the threshold of legal precedence and public confidence.

ACKNOWLEDGEMENTS

The authors wish to thank Don Tompkins and Drs. R. Aubin and J.H. Bowen for their helpful comments on the manuscript.

REFERENCES

Adams, D. E. (1988). Validation of the FBI procedure for DNA analysis: A summary, Crime Lab. Digest 15:106–108.

Adams, R. L. and Burdon, R. H. (1985). Methylation in higher eukaryotes, In: Molecular Biology of DNA Methylation, Rich, A. ed., Springer Verlag, NY, 89–114.

Blakesley, R. W., Dodgson, J. B., Nes, I. F. and Wells, R. D. (1977). Duplex regions in single-standed phi X174 DNA are cleaved by a restriction endonuclease from Haemophilus aegyptius, J. Biol. Chem. 252:7300–7306.

Botstein, D., White, R., Skolnick, M. and Davis, R. (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms, Amer. J. Hum. Genet. 32:314–331.

Bowtell, D. D. L. (1987). Rapid isolation of eukaryotic DNA, Anal. Bioch. 162:463–465.

Budowle, B., Deadman, H. A., Murch, R. S. and Baechtel, F. S. (1988). An introduction to the methods of DNA analysis under investigation in the FBI Laboratory, Crime Lab. Digest. 15:8–21.

Budowle, B., Waye, J. S., Shutler, G. G. and Baechtel, F. S. (1990). HaeIII A suitable restriction endonuclease for restriction fragment length polymorphism analysis of biological evidence samples, J. Foren. Sci. 53: (in press).

Burke, T. and Bruford, M. W. (1987). DNA fingerprinting in birds, Nature 327:149–152.

Cory, S., Gerondakis, S. and Adams, J. M. (1983). Interchromosomal recombination of the cellular oncogene c-myc with the immunoglobulin heavy chain locus in murine plasmacytomasis a reciprocal change, EMBO 2:697–703.

Fourney, R. M., Dietrich, K. D. and Paterson, M. C. (1989). Rapid DNA extraction and sensitive alkaline blotting protocol: Application for detection of gene rearrangement and amplification for clinical molecular diagnosis, Disease Markers 7:15–26.

Gill, P., Jeffreys, A. J. and Werrett, D. J. (1985). Forensic application of DNA "fingerprints", Nature 318:577–579.

George, M., Lequarre, A.-S., Castelli, R. and Vassart, G. (1988). DNA fingerprinting in domestic animals using four different minisatellite probes, Cytogenet. Cell Genet. 47:127–131.

Gross-Bellard, M., Oudet, P. and Chambon P. (1973). Isolation of high-molecular-weight DNA from mammalian cells, Eur. J. Biochem. 36:32–38.

Higuchi, R., von Beroldingen, C. H., Sensabaugh, G. F. and Erlich, H. A. (1988). DNA typing from single hairs, Nature 332:543–546.

Hill, W. G. (1987). DNA fingerprinting applied to animal and bird populations, Nature 327:98–99.

Huang, L. H., Farnet, C. M., Ehrlich, K. C. and Ehrlich, M. (1982). Digestion of highly modified bacteriophage DNA by restriction endonucleases, Nucleic Acids Res. 10:1579.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985a). Hypervariable "minisatellite" regions in human DNA, Nature 314:67–73.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985b). Individual-specific "fingerprints" of human DNA, Nature 316:76–79.

Jeffreys, A. J., Wilson, V., Thein, S. L., Weatherall, D. J. and Ponder, B. A. J. (1986). DNA "fingerprints" and segregation analysis of multiple markers in human pedigrees, Am. J. Hum. Genet. 39:11–24.

Jeffreys, A. J. and Morton, D. B. (1987). DNA fingerprints of dogs and cats, Animal Genetics 18:1–15.

Law, D. J., Frossared, P. M. and Rucknagel, D. L. (1984). Highly sensitive and rapid gene mapping using miniaturized blot hybridization: application to prenatal diagnosis, Gene. 28:153–158.

Lerman, L. (1986). Introduction. In: DNA Probes Applications in Genetic and Infectious Disease

and Cancer. (Lerman, L. S. ed.), Current Communications in Molecular Biology, Cold Spring Harbor Laboratory, NY, 1–5.

Longmire, J. L., Albright, K. L., Lewis, A. K., Meincke, L. J. and Hildebrand, C. E. (1987). A rapid and simple method for the isolation of high molecular weight cellular and chromosome specific DNA in solution without the use of organic solvents, Nucleic Acids Res. 15:859.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). In: Molecular Cloning: A Laboratory Manual, Cold Spring Harbour Laboratory, Cold Spring Harbour, NY.

McClelland, M. and Nelson, M. (1988). The effect of site-specific methylation on restriction endonucleases and DNA modification methyltransferases, Gene. 74:291–301.

Monson, K. L. (1988). Semiautomated analysis of DNA autoradiograms, Crime Lab. Digest. 15:104–105.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Mastin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987a). Variable number of tandem repeat (VNTR) markers for human gene mapping, Science 235:1616–1622.

Nakamura, Y., Gillilan, S., O'Connell, P., Leppert, M., Lathrop, G. M., Lalouel, J.M. and White, R. (1987b). Isolation and mapping of a polymorphic DNA sequence pYNH24 on chromosome 2 (D2S44), Nucleic Acids Res. 15:10073.

Reed, K. C. and Mann, D. A. (1985). Rapid transfer of DNA from agarose gels to nylon membranes, Nucleic Acids Res. 13:7207–7221.

Reymond, C. D. (1987). A rapid method for the preparation of multiple samples of eucaryotic DNA, Nucleic Acids Res. 15:8118.

Ryskov, A. P., Jincharadze, A. G., Prosnyak, M. I., Ivanov, P. L. and Limborskaya, S. A. (1988). M13 phage DNA as a universal marker for DNA fingerprinting of animals, plants and microorganisms, FEBS Letters 233:388–392.

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. I., Erlich, H. A. and Arnheim, H. (1985). Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia, Science 230:1350–1354.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, Science 239:487–491.

Silva, A. J. and White, R. (1988). Inheritance of allelic blueprints for methylation patterns, Cell 54:145–152.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis, J. Mol. Biol. 98:503–517.

Waye, J. S., England, S. B. and Willard, H. F. (1987). Genomic organization of alpha satellite DNA on human chromosome 7: Evidence for two distinct alphoid domains on a single chromosome, Mol. Cell Biol. 7:349–356.

Waye, J. S., Presley, L. A., Budowle, B., Shutler, G. and Fourney, R. M. (1989). A simple and sensitive method for quantifying human genomic DNA in forensic specimen extracts. (Biotechniques, in press).

Waye, J. S. and Fourney, R. M. (1989). Agarose gel electrophoresis of linear genomic DNA in the presence of ethidium bromide: Band shifting and implications for forensic identity testing (Appl. Theoret. Electro. in press).

Weeks, D. P., Beerman, N. and Griffith, O. M. (1986). A small-scale five-hour procedure for isolating multiple samples of CsCl-purified DNA: Application to isolations from mammalian, insect, higher plant, algal, yeast and bacterial sources, Anal. Bioch. 152:376–385.

Weller, P., Jeffreys, A. J., Wilson, V. and Blanchetot, A. (1984). Organization of the human myoglobin gene, EMBO. J. 3:439–446.

Welton, J. H., Carter, R. E., Parklin, D. T. and Walters, D. (1987). Demographic study of a wild house sparrow population by DNA fingerprinting, Nature 327:147–152.

Werrett, D. J. (1987). DNA fingerprinting, Int. Crim. Police. Rev. Sept - Oct:21–25.

White, T. J., Arnheim, N. and Erlich, H. A. (1989). The polymerase chain reaction, TIGS 5:185–189.

Wyman, A. R. and White, R. (1980). A highly polymorphic locus in human DNA, Proc. Natl. Acad. Sci. USA 77:6754–6758.

DNA ANALYSIS IN HOME OFFICE LABORATORIES
ITS INTRODUCTION, IMMEDIATE FUTURE AND STATISTICAL ASSESSMENT

David J. Werrett, Peter D. Gill, Ian W. Evett, J. E. Lygo, K. M. Sullivan and J. Buckleton

Home Office Forensic Science Service
Central Research and Support Establishment
Aldermaston, United Kingdom

In January 1987 the Home Office Forensic Science Service (HOFSS) for England and Wales began a casework in DNA analysis. A probe, Ig51s (Werrett *et al.* 1988), containing 25 repetitive elements homologous to the probe 33.15 described by Jeffreys *et al.* (1985a,b), was used to obtain multilocus profiles from case stains (Figure 1). Initially cases were carefully selected and only those of a serious nature were examined. Experience has shown that multilocus probe (MLP) analysis offers the advantage of providing discriminating information from one test that can be of the order of 1 in millions (Gill *et al.* 1987). However the sensitivity

Figure 1. Multilocus profiles from case stains: A = accused, D = deceased, C = control. Note the match between the suspects sample A (blood sample result on the left and bloodstain result on the right) and the profile obtained from the vaginal swab. The vaginal swab and the suspect share 10 bands in common above 4 Kb: chance association less than 1 in 700,000. The faint profiles obtained from the anorak also contain bands of similar electrophoretic mobility but note the overall slight shift of the pattern. the single locus result, MS1, also shows this shift. The samples were run in 0.7% agarose gels using recirculated TAE buffer until the 2.0 Kb of a lambda HindIII digest had moved approximately 18 cm.

of the test (approximately 0.5 ug of DNA being required for a full profile), limits its application particularly in those cases where bloodstain analysis is required (Werrett and Lygo 1989).

Since January 1987 the HOFSS has trained staff from each of the six operational laboratories in DNA analysis and has provided purpose built DNA units at each of these sites. Several hundred cases have been examined and about 50 of these have been through the courts (Table 1). The technique as performed by each of these units, was assessed initially by a validation study before casework was carried out. The work is now monitored by:

Table 1. CASES THAT HAVE BEEN PRESENTED TO THE COURTS

Type of offence	No. Cases
Rape (Att. Rape)	37
Murder	6
Incest	3
Assault	1
Disputed Maternity	1

Plea	Verdict	Number
Not Guilty	Not Guilty	11
Not Guilty	Guilty	19
Guilty	Guilty	17
Not Guilty	Retrial	1

Details of Not Guilty Verdict Cases	
Number	Details
4	The accused claimed the victim consented
3	DNA evidence favoured the defence
2	Complainant was unable to continue
1	Semen on bed sheet, no vaginal swab, rape considered not proven
1	SLP analysis showed accused seminal DNA was a minor component

(i) Declared quality assurance trials
(ii) Undeclared quality assurance trials

(iii) Revision of the band sharing statistic 'x', that is, the probability that a band in individual A will also be present in a second individual B

(iv) The DNA Working Group

The validation study comprised 50 samples: 20 blood samples, 12 bloodstains, 12 semen stains and 6 vaginal swabs (some or all of which were said to be semen stained). All 6 operational laboratories successfully completed this study which required that the donors of the blood and semen stains be identified by comparison with the control blood samples. Some laboratories failed to obtain a result for all of the stains submitted but there were no mis-identifications (Table 2). The study allowed the comparison of results both within and between laboratories (Figures 2a and 2b).

Table 2. THE VALIDATION OF THE SIX HOME OFFICE LABORATORIES

Laboratory	Blood Sample	Blood-Stain	Semen Stain	Vaginal Swab*	Total Results	No. Results
1	20	12	9	3	44	6
2	19	12	12	3	46	4
3	20	12	11	3	46	4
4	19	12	12	4	47	3
5	20	12	8	4	44	6
6	19	12	12	4	47	3

The numbers presented in each column are the number of samples that were successfully profiled by each laboratory in the study: there were 20 blood samples, 12 bloodstains, 12 semen stains and 6 vaginal swabs. All of the donors of the blood and semen stains (including the semen on the vaginal swabs) were correctly identified, with reference to the control blood samples, where results were obtained.

* Only 4 of the 6 vaginal swabs were semen stained.
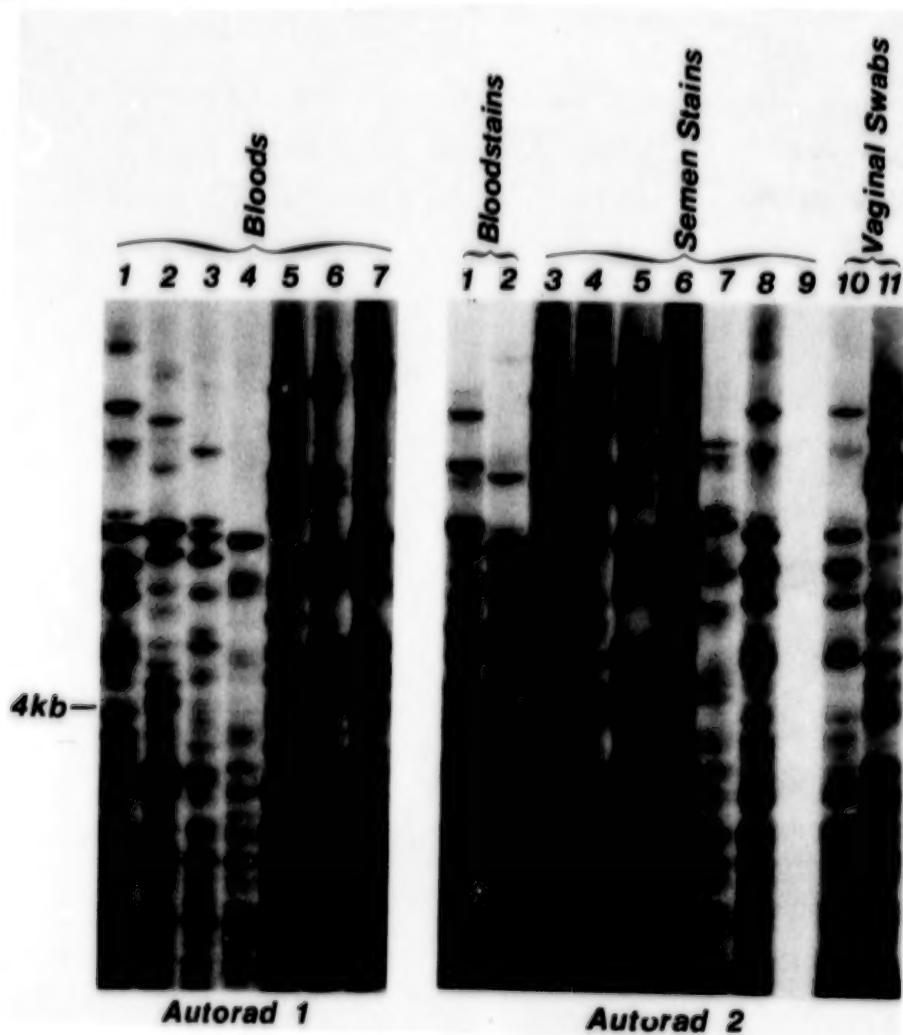
Figures 2a and 2b. Results obtained from the validation studies. These studies allowed the comparison of results both within (Figure 2a) and between laboratories (Figure 2b). In Figure 2a (above) sections of 2 autoradiographs from one laboratory are presented, matching samples are seen in lanes 1 (liquid blood) autoradiograph 1 and lanes 8 (semen stain) and 10 (vaginal swab) autoradiograph 2, also lane 7 (liquid blood) autoradiograph 1 and lane 5 (semen stain) autoradiograph 2 match. Further within autoradiograph 2 lane 4 (semen stain) matches lane 11 (vaginal swab).

Quality assurance is an essential part of forensic analysis, the Quality Assurance (QA) Division of the HOFSS provides both declared and undeclared trials. The declared trials comprise bimonthly submissions of 3 blood samples and 3 semen stains, some of the samples may match each other (Figure 3). The undeclared trials take the form of mock cases, or part of a case, which are submitted to the laboratories, without their knowledge, via the police. In the case which involved a screen of more than 5,000 samples, because two teenage girls had been sexually assaulted and murdered in the vicinity of three small villages (Gill and Werrett 1987), QA Division monitored the efficacy of the screen by adding 6 undeclared samples to it. These were correctly eliminated and the expected results obtained.
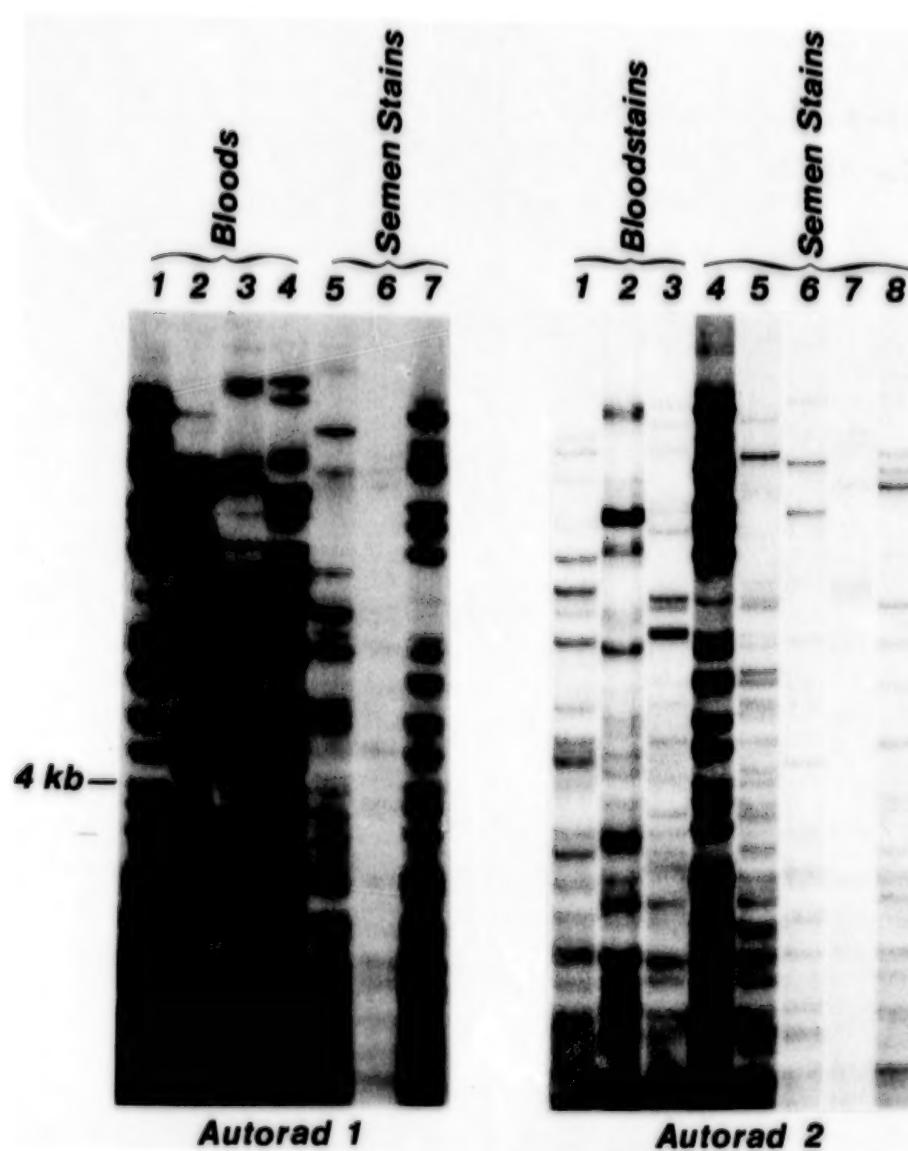
In Figure 2b the autoradiographs 1 and 2 were submitted by different laboratories, matching samples are seen in lanes 1 (liquid blood) and 7 (semen stain) autoradiograph 1 and these match lane 4 (semen stain) on autoradiograph 2. Also lane 6 (semen stain) autoradiograph 1 matches lane 8 (semen stain) autoradiograph 2. It is important with multilocus analysis that a common protocol is carefully followed by each laboratory. Changes to the protocol, particularly with regard to hybridization conditions, could effect the profiles obtained and may also effect the degree of band sharing between random individuals. The figure of 0.26 used by HOFSS laboratories is considered to be conservative.
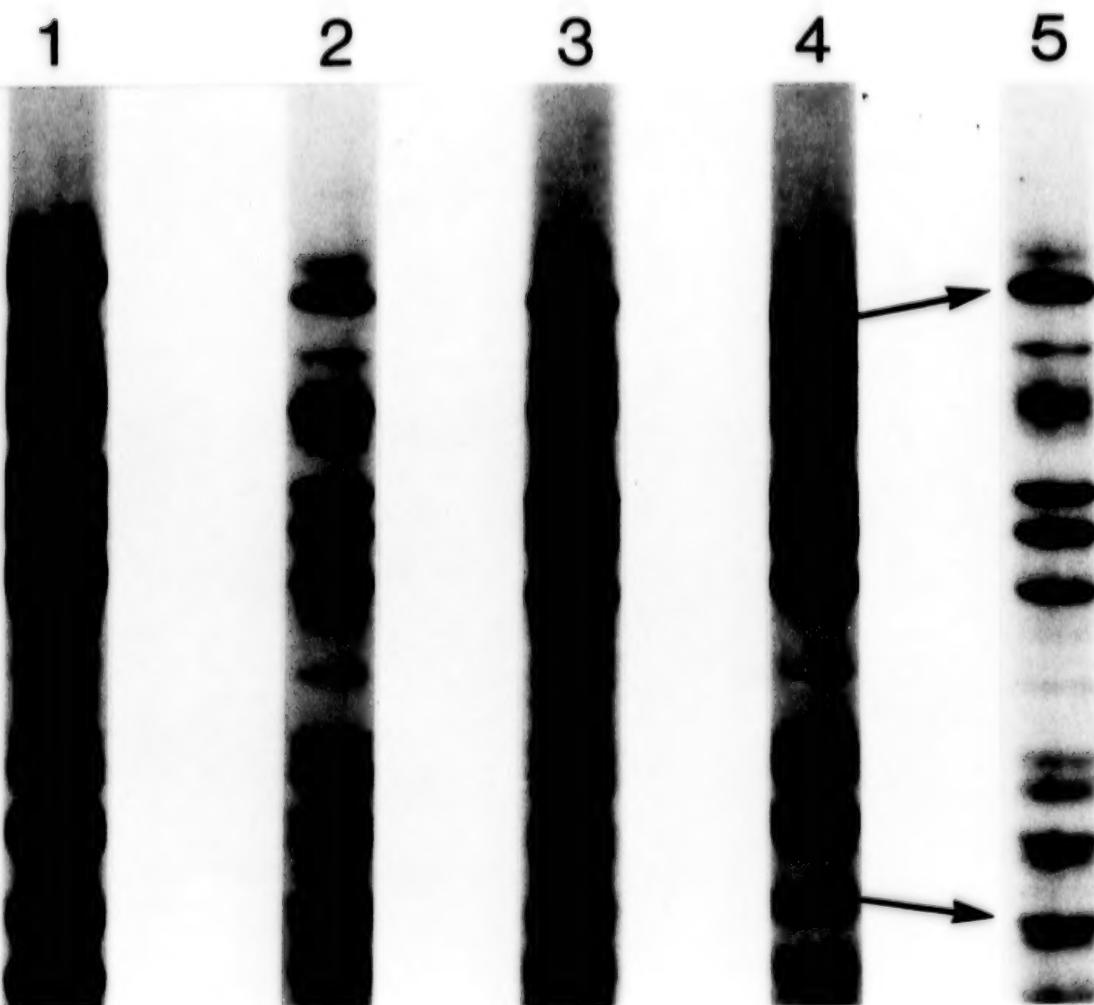
Figure 3. Results submitted by five of the HOFSS laboratories as part of the first QA Trial. The sixth operational unit had not at this stage been validated. Note how the patterns are reproducible, however, the over exposure of the autoradiograph from laboratory 1 has made the bands less distinct. For comparative purposes the profiles would require normalization due to the slightly different running conditions, for example, laboratory 5 has run the sample further than laboratory 4: equivalent band positions are indicated by the adjoining lines. These early declared trials have also demonstrated the importance of ensuring complete restriction and the need for the careful interpretation of the profiles.

Revision of the band sharing statistic 'x', that is, the probability that a band in individual A will also be present in a second individual B. The calculation of chance association of multilocus profiles has been based on the statistic: $1/x^j$ where j is the number of bands shared between the two profiles being compared. The statistic may be thought of as a truncated binomial expression, the full binomial being: $1/x^j(1-x)^{N-j}$ where $(1-x)$ is the probability that a band is not present and N the number of positions at which a band may present. The statistic $1/$

x^j is derived empirically by comparison of large numbers of profiles with a single control sample as described by Jeffreys et al. (1985a,b) and Gill et al. (1987). When making the comparison, any band that comigrates with one chosen from the control sample is scored as a matching band. Clearly in casework analysis matching bands must also be of the same relative intensity. The conservative figure of 0.26 has been chosen, this exceeds the estimate of band matching found for any of the size classes 4–6, 6–9 and 9-23 kilobase (Kb) (Table 3).

Table 3. BAND SHARING STATISTICS FOR THE SIX HOME OFFICE LÅBORATORIES

Lab	Number Profiles	23–9KB Total Shared		9–6KB Total Shared		6–4KB Total Shared		Overall Total Shared	
1	130	390	35	397	68	689	173	1476	275
		(0.14)		(0.17)		(0.22)		(0.19)	
2	272	807	49	731	155	1404	325	2946	529
		(0.09)		(0.19)		(0.20)		(0.18)	
3	96	357	22	311	65	481	142	1149	229
		(0.12)		(0.23)		(0.25)		(0.22)	
4	101	367	26	320	54	639	134	1326	214
		(0.13)		(0.18)		(0.22)		(0.19)	
5	91	344	22	301	52	555	122	1200	196
		(0.12)		(0.19)		(0.22)		(0.20)	
6	50	116	5	124	19	267	74	509	98
		(0.05)		(0.13)		(0.25)		(0.18)	
Overall Figures	**740**	**2381**	**159**	**2184**	**413**	**4037**	**970**	**8606**	**1541**
Overall band sharing		0.11		0.19		0.22		0.19	
Average number		3.22		2.95		5.46		11.63	
Bands SD		1.64		1.20		1.52		2.75	

The total figure is provided to give the reader an overall assessment of band sharing obtained by the 6 laboratories but it should also be noted that differences were obtained for the band sharing statistics between the laboratories (figures in parenthesis). All profiles were assessed by the same two operators. Similar results have been obtained from other smaller exercises carried out using different operators and different control samples. Overall the adoption of the value 0.26 for an assumed constant band sharing statistic x over the size classes 4 to 34KB is considered to be conservative. True values of x clearly vary between and by inference probably within size classes but do not exceed the value of 26.

The DNA Working Group comprises one individual from each of the Home Office Laboratories and a representative from the Metropolitan Police Laboratory. It is chaired by the Head of Biology, CRSE. The group reports to the Director of Research CRSE who in turn reports to the Director of General of the HOFSS. The aims and objectives of the group are to advance DNA analysis and its application in forensic investigation and to coordinate approved changes. Currently the members of the group are assessing several single locus probes (SLPs) for casework application: the probes are pCRE 1.2, pCRE 2.7, derived from probes pMLJ14 and YNH24 respectively (Nakamura *et al.* 1987), 3' HVR (Jarmen *et al.* 1986) and MS43A (supplied by Cellmark Diagnostics, UK). The probes pMLJ14 and YNH24 were gratefully received from Y. Nakamura of the University of Utah, Salt Lake City, Utah. In their original form these probes, particularly pMLJ14, gave relatively high backgrounds. In the case of pMLJ14 the original 40 Kb cosmid insert was digested with Msp1 and the fragments oligolabelled. A group of three fragments were found to identify the hypervariable region and these were shotgun cloned with a Boehringer cloning kit. From the resultant colonies a single colony was found to identify the same restriction fragment length polymorphisms (RFLPs) as pMLJ14 with minimal background hybridization (Figure 4) and this was designated pCRE1.2.YNH24 has been similarly subcloned to form pCRE 2.7.
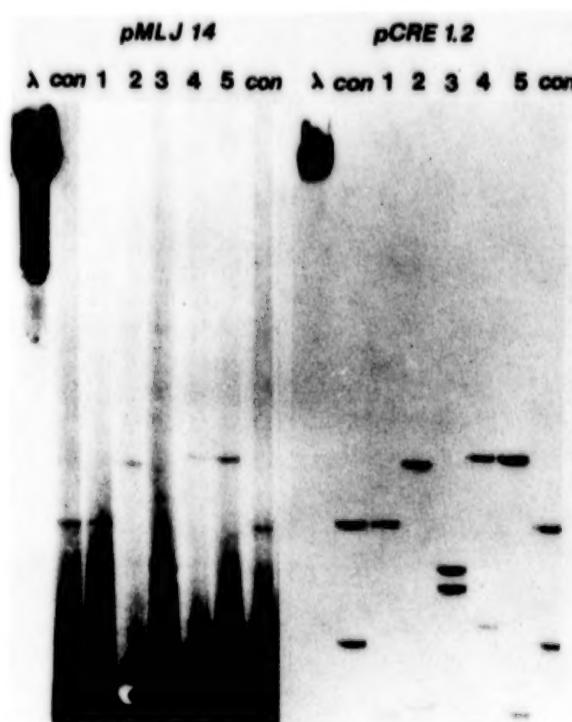

Figure 4. Autoradiographs of genomic DNA digested with Hinfl and hybridized with pMLJ14 on the left and pCRE1.2 on the right, pCRE1.2 contains only the 600 base pair Msp 1 fragment of pMLJ14 and detects the same RFLPs, it is excised from the plasmid pUC 18 with BamHI and Hind III. The clearer backgrounds obtained with pCRE1.2 allows the elucidation of the weakly hybridizing lower molecular weight alleles.

It is envisaged that MLPs and SLPs will be used in tandem, although it is recognized that it

may not always be possible to combine the statistics derived from these forms of DNA analysis. Some of the advantages gained from using both methods of analysis, multilocus and single locus, are demonstrated in Figures 5a and 5b. The techniques complement one another. In Figure 5 the preferentially extracted semen (Gill *et al.* 1985) is checked for contaminating male DNA by the single locus analysis. While in Figure 5b a putative match between the suspect's sample and semen staining on two items is confirmed by the multilocus approach. Distortion in tracks, as shown by Figure 5b and to a lesser extent Figure 1, has led to a cautious approach with regard to the adoption of SLPs.

Apart from an increase in sensitivity it is considered that SLPs will provide a more convenient basis for interlaboratory comparison of results, par-
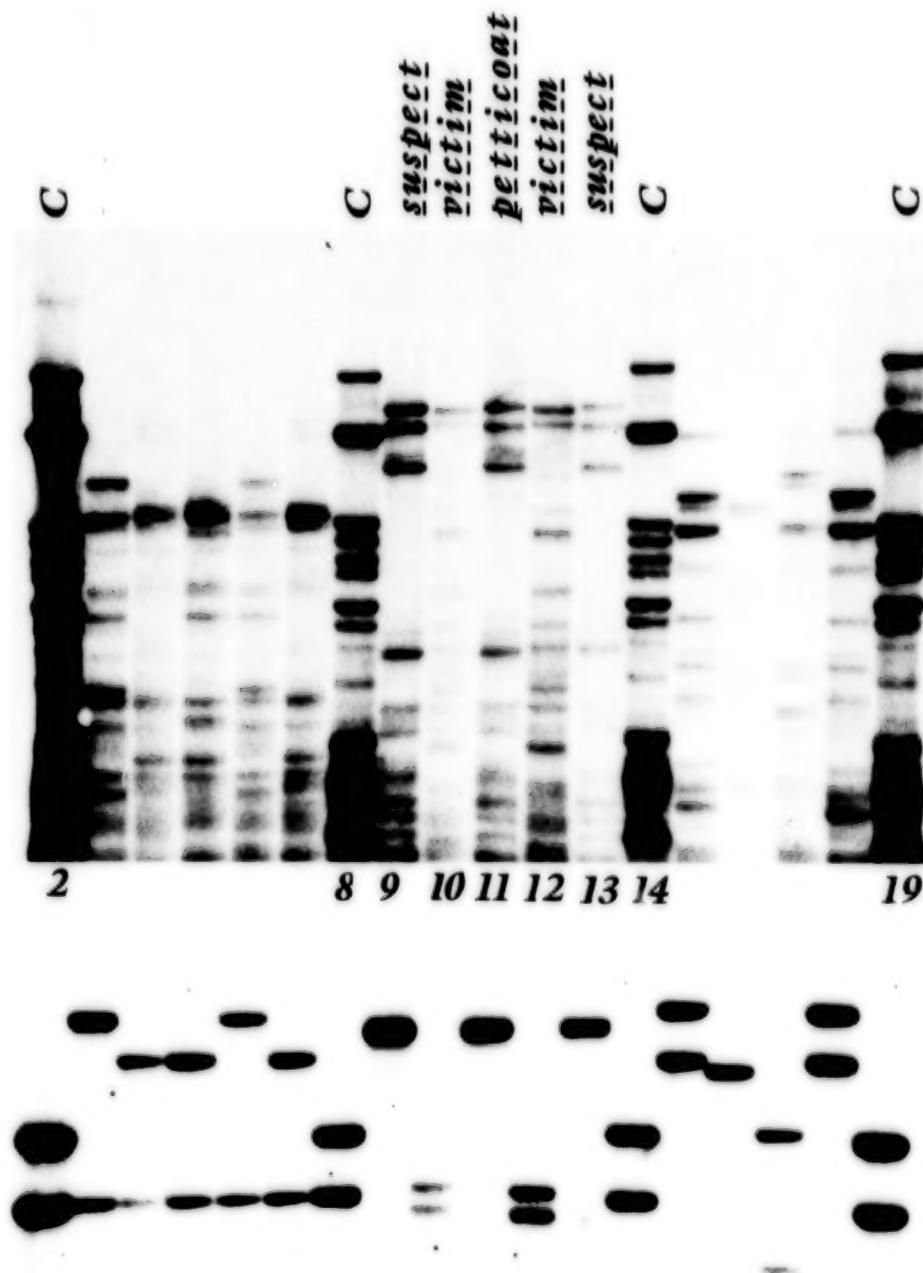
Figures 5a. In Figure 5a the multilocus result from the preferential extraction technique is checked for contaminating female DNA by subsequent single locus analysis: the high molecular weight bands of both the victim (tracks 10 & 12) and suspect (tracks 9 & 13) are similar in electrophoretic mobility, however the single locus result (overexposed for the purpose of the test) confirms the absence of detectable contaminating female DNA from the petticoat Track 11.

ticularly when the laboratories do not belong to the same organization, or even country (In Europe 12 laboratories are collaborating in an interlaboratory comparative study using pCRE2.7 (YNH24) as the probe of choice). The results obtained from MLPs are sensitive to the precise control of the low stringency hybridization conditions. This may have little effect on the overall band sharing statistic x but it could make interlaboratory comparisons very difficult. Essentially each laboratory must closely follow the recommended protocol. As an added precaution, to confirm the match, blots may be stripped and re-hybridized in the same chamber. The Home Office laboratories work in close cooperation, as they are part of the same organization. However, it is doubtful whether such collaboration could be achieved across organizational boundaries. The use of multilocus analysis for DNA data banks is, therefore, considered questionable and so selected SLPs are being assessed.

The assessment of these probes may be subdivided:

(i) Development of the probes, as required, at the CRSE (as previously described)

(ii) The analysis of at least 200 samples from the three major ethnic groups: White caucasians, Afro-carribeans and Indian/Pakistanis

(iii) The construction of data banks on Biotrac (Foster & Freeman, Evesham, UK)

(iv) The derivation of a method of statistical assessment.

Stages (i) and (ii) are virtually complete but the problems encountered at stage (iii) have led to the development of a novel method of statistical assessment at the CRSE.

Superficially it would appear that the interpretation of single locus results should mirror current conventional analysis. The blood sample from the suspect may be tested together with the stain from the scene of crime and if a match is obtained the frequency of joint occurrence of the two alleles quoted, that is, 2pq. However, we have found both the definition of an allele and a match to be problematical for hypervariable loci. The analysis of hypervariable loci pose both of these questions because firstly the bands identified are anonymous, they cannot be named as for example the alleles of the PGM1 locus are: 2+, 2−, 1+, 1−. Secondly for some probes the distribution of bands is virtually continuous and closely migrating bands may represent the same anonymous allele or two different alleles of similar electrophoretic mobility.

It is essential, therefore, that the analyst considers the alternatives, that is, the control and
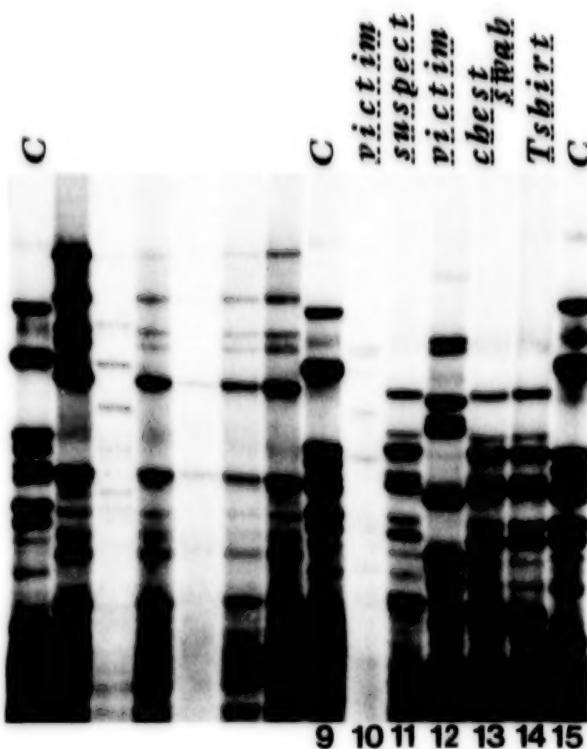
Figure 5b: The distortion of the plate makes single locus comparison difficult of Tracks 11, 13 & 14: the sample from the suspect (track 11) shows two bands that have similar electrophoretic mobility to the semen staining on the chest swab (track 13) and T shirt (track 14) but there remains the possibility that the suspect has two closely migrating but different alleles at this locus (D2S44). The latter is shown to be almost certainly incorrect by the multilocus result.

These results were kindly supplied by Miss Linda Nelson of the Home Office Forensic Science Laboratory at Aldermaston.

crime samples are of common origin or the control and crime samples contain closely migrating but different alleles. An assessment will need to be made as to which of these two alternatives is more

likely. If the former is the case then the analyst will also need to enumerate the findings in some way so as to weigh the value of the evidence. In an attempt to resolve the problem a simple model was used: a Lambda ladder from BRL (Bethesda Research Laboratories, Gaithersburg, MD) containing 12 fragments of known molecular weight (range 1.6 to 12.2 Kb). The ladder was analyzed 42 times using three gels (Figure 6). The object of the analysis was to explore the derivation of frequency of band occurrence in a multiallelic system and to identify a method of describing a band match, that is, a means of stating that bands A and B had, or could have had, a common origin. In the simple model chosen, the replicate analysis of a molecular weight marker, the test is particular uninformative and also confounds some basic principles of genetics as each band has a frequency of occurrence of 100%! But we do know that each sample is potentially a perfect match for every sample. Thus our analysis of these profiles should demonstrate a 100% occurrence of each allele and also that each sample matches every other sample.

The Biotrac machine was used to analyze the data. This machine assigns bands to 600 arbitrary positions on a linear scale, comparison between autographs is facilitated by reference to control lanes containing bands of known size. The disadvantage of this process is that the linear extrapolations between controls ignores any distortions within a track and this may give rise to an inaccurate assessment of band position. Required band position can be subsequently converted to molecular weight by polynomial fitting or by the reciprocal method of Elder and Southern (1987). Analysis of the 42 samples generated band distributions could provide erroneous frequencies of occurrence if no account was taken of the errors incurred in the analysis (Figure 6). If read directly from the histogram, none of the frequencies of occurrence exceeded 40%, greatly underestimating the true 100% frequency of occurrence of the alleles (Figure 7). For individual bands the standard deviation (SD) was approximately 2, range from 1.75 to 2.36 position units where 1 unit = approx. 0.38 mm (Table 4). Of note is that with respect to molecular weight the error in determining band size varies, thus the conversion of SD to molecular weight for the 12.2 Kb fragment gives approximately ± 227 bases whereas for the 1.6 and 2.0 Kb fragment the figure is approximately ± 20 bases.

Previous workers have determined allele frequencies by binning (Baird *et al.* 1986) This procedure entails dividing an autoradiograph into a number of sectors which are based on 95% confidence limits (8 position units in this example). Application of this procedure to the model used here has two undesirable effects when the allele distribution does not fall discretely into one bin: the frequency of occurrence of the band is underestimated and because of the sub-division, in this particular instance, 20 alleles are recorded instead of 12 (Figure 8). We realize that this is a severe test of the binning procedure however the distribution of alleles at hypervariable loci does tend to be spiked with relatively common alleles (Wong *et al.* 1986, 1987). In practice the frequency of these more common alleles will be susceptible to distortion by the binning procedure.

An alternative but very conservative method of frequency determination employs a sliding window routine. The routine takes each of the 600 positions in turn and sums the frequency of occurrence of bands within ± 2 SD of that point, essentially the treatment assumes that each point is the true position of any given allele. Using this method on the model correctly identifies 12 distinct peaks (Figure 9). However the true frequency of occurrence will only be obtained when the band falls in the middle of the distribution. When the bands fall into the tail of the distribution another routine, for example, the scanning of the histogram for the maximum frequency within ± 3 SDs of the data point, will be required to ensure that the true frequency of occurrence is obtained. For hypervariable loci it is almost certain that estimates of frequency of occurrence derived by this method will be highly conservative. Considering the area under the frequency curve, it will be equivalent to 1 + 4 SD or 900% (1 SD was equivalent to approximately 2 position units). As an analogy consider the effect on the frequencies of the ABO system if collectively they were adjusted to add up to 900%!

Both the sliding window routine and binning provide little assistance to the operator in deciding when bands match. From the model it is possible for bands to be more than 9 position units apart yet still be a match, however the probability of two matching bands falling into opposite tails of the allele's error distribution must be far less than the bands falling for example within one standard deviation of the mean. It is, therefore, essential that the statistic quoted in some way evaluates that probability.

From the above it is clear that both the assessment of band matching and frequency of occurrence require another approach. Furthermore, as we have seen, for hypervariable loci it is essential
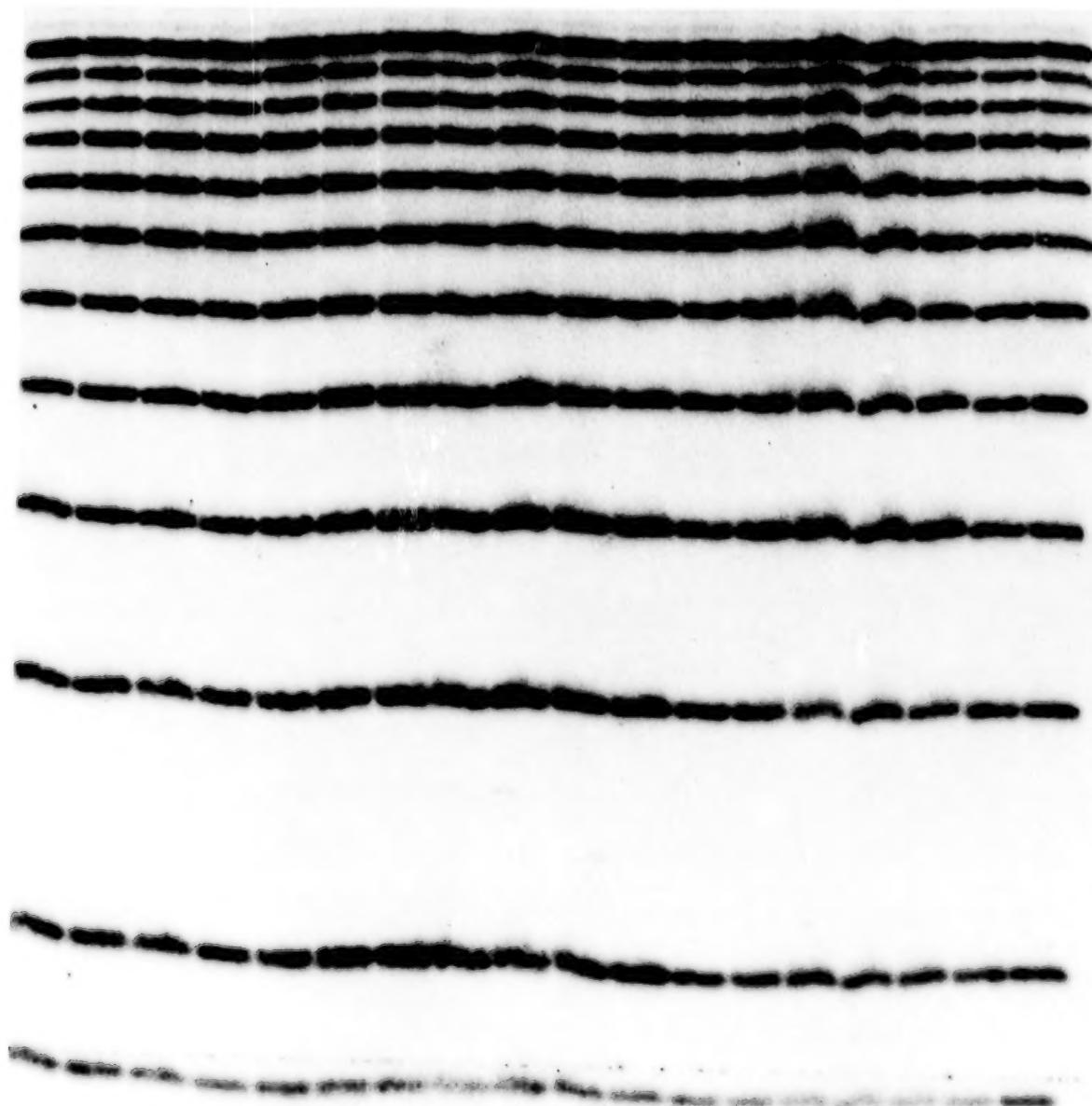
Figure 6. The analysis of a molecular weight ladder: lambda ladder from **BRL** containing 12 fragments in the molecular weight range 1.6 to 12.2 **Kb**. Note the fluctuation of band position. The ladder was used to explore the ability to arrive at the correct frequency of occurrence for alleles (in this artificial case each allele has a 100% occurrence!) and the requirements for matching of bands. The ladder has been run in a 0.6%, 30 cm agarose gel using Tris-borate buffer.

that the analyst, when considering a possible match, considers the alternatives, that is, the control and crime samples are of common origin or the control and crime samples have come from different people. This statement may generalized to:

The stain was left by the suspect

The stain was left by someone else

Consideration of the results in this manner falls conveniently into a Bayesian type analysis where the alternatives that the forensic scientist has to

consider form the likelihood ratio of the Bayesian approach:

Posterior Odds = Likelihood ratio x Prior Odds

In terms of the court the posterior odds are concerned with the probability of guilt/innocence given the evidence as a whole, the likelihood ratio is the probability of the forensic evidence given the suspect either did or not leave the stain and the prior odds are the probability of guilt/innocence
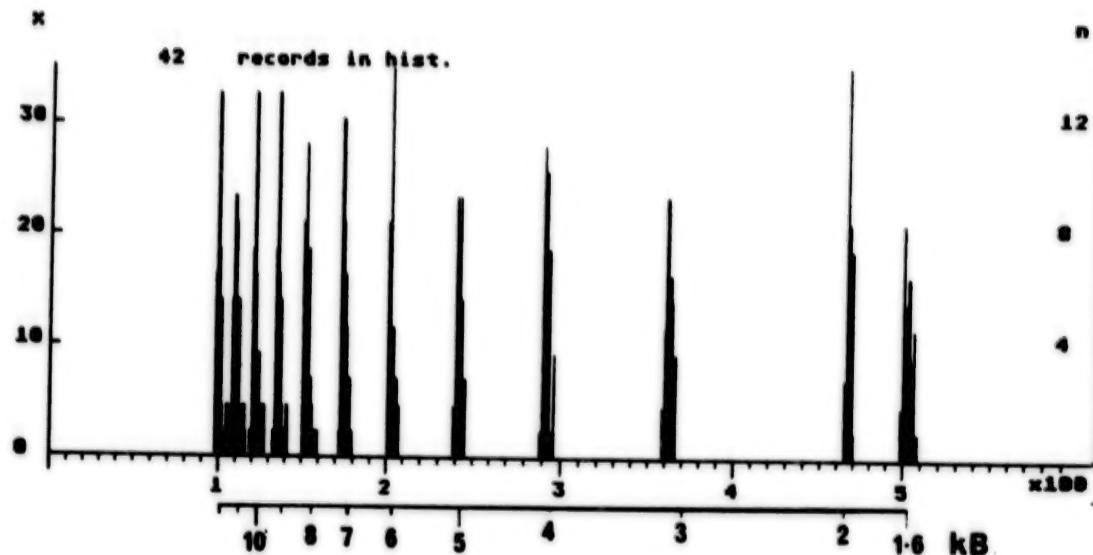
Figure 7. The analysis of the molecular weight ladder: the peak of the frequency of occurrence is less than 40%: greatly underestimating the true frequency of occurrence, that is, 100%.
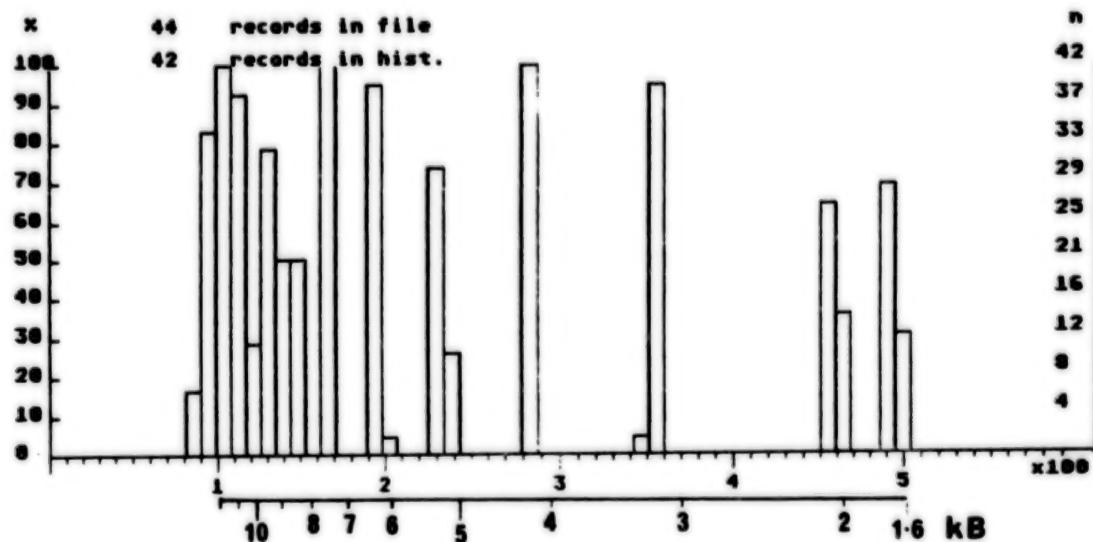
Figure 8. Histogram produced by binning the bands based on + - two standard deviations, the bins are therefore 4 SDs wide. Note the that 20 alleles have now been recorded instead of 12 and that where the bin subdivides the distribution of a band the true frequency of occurrence is underestimated.

before evidence was presented. The forensic scientist is concerned with the likelihood ration, that is,

The probability of the evidence given that the stain was left by the suspect

The probability of the evidence given that the stain was left by someone else

If the likelihood ration is greater than one it will tend to support the hypothesis that the control and crime stain have common origin whereas values of less than one will support the hypothesis that the stain has orig nated from someone else.

This form of analysis follows in the spirit of that originally used by Lindley (1977) who demon-

strated that the two stage approach (it matches, what is the frequency of occurrence?) was unacceptable for glass analysis. The Bayesian approach to the evaluation of forensic evidence has been advocated by others: Evett *et al.* (1987) addressed the value of fiber evidence in the Bayesian framework and Hill (1986) was the first to advocate the use of a Bayesian analysis for the assessment of parentage based upon MLP profiles. More recently Gjertson *et al.* (1989) have described a detailed Bayesian analysis for hypervariable SLPs. However, Gjertson *et al.* (1989) used the method of binning to arrive at allele distributions, the artificial assump-

156
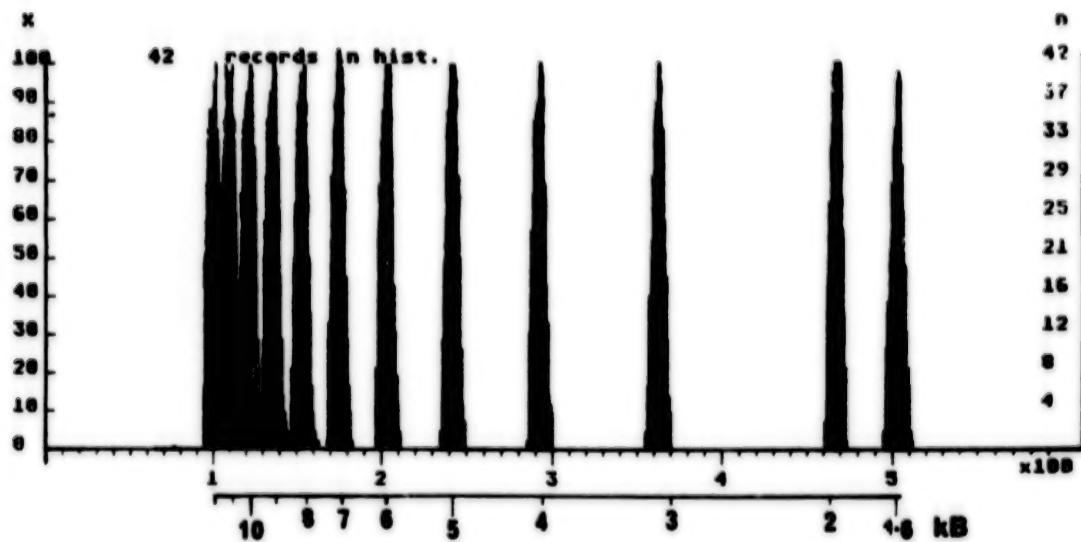
Figure 9. A sliding window fit using 2 SDs. Each point represents the sum of the occurrence of bands at that point plus the sum of bands occurring within + 2 SDs. The method correctly identifies 12 peaks at 100% occurrence. However bands falling into the tail of the distribution will be given under estimates of their frequency of occurrence unless a further routine is employed to identify peaks within the vicinity of the band position.

tion was made that each interval in the histogram (a bin) represented an allele. It is possible, within the Bayesian framework, to circumvent the problems encountered with attempts to define allele distributions.

We have established above that the observed distribution of bands for a hypervariable locus is a mixture of two separate distributions:

1. A discrete allele size distribution
2. A continuous measurement error distribution appears to be normally distributed (Figure 10) and approximately constant along the arbitrary linear scale used by Biotrac (Table 4).

The effect of the measurement error will be to smooth out the underlying discrete distribution. For the derivation of the likelihood ratio we can use this convoluted distribution. For two banded individuals, assuming a randomly mating population, the measurement of each band can effectively be treated as two independent samples from the same distribution.

Let Z denote the vector of two measurements made on the two bands of a person selected at random from the population. We require the probability distribution $f(Z = z)$, or $f(z)$. To obtain this distribution a sample of the population is required: for the HOFSS 200 individuals of each of the three main ethnic groups which will provide up to 2×200 observations. These $\sim 2n$ observations will represent an estimate of the true probability distribution. The distribution will contain gaps which are likely to be filled as more samples are analyzed (Figure 11). This is to be expected from our exami-

nation of binned allele data it is apparent that the number of alleles recorded is related to the size of the population examined.

Table 4. RESULTS OBTAINED FROM THE ANALYSIS OF THE BRL LADDER

Modecular Weight	SD (Position)	Range	SD (Bases)	Range (Observed)
12.2	2.16	8	227(1.9%)	840(6.8%)
11.2	2.05	8	197(1.8%)	770(6.9%)
10.2	1.96	8	152(1.5%)	620(6.1%)
9.2	1.87	8	110(1.3%)	450(4.9%)
8.1	1.84	8	110(1.3%)	480(5.9%)
7.1	1.75	7	70(1.0%)	280(3.9%)
6.1	1.76	6	70(1.2%)	240(3.9%)
5.1	1.58	6	40(0.8%)	130(3.2%)
4.1	1.92	7	36(0.8%)	130(3.2%)
3.05	1.9	7	19(0.6%)	70(2.3%)
2.0	1.34	5	16(0.8%)	60(3%)
1.6	2.36	9	23(1.4%)	80(5%)
Mean	1.95	7.2	(1.25%)	

Note the relationship between range and molecular weight, as a consequence smaller alleles are more accurately sized. One positional unit represents 0.375mm.

Thus from the samples band distribution we need to estimate the true population distribution of $f(z)$. This can be done by smoothing the data and we have used as a smoothing function the measurement error distribution. Thus if we assume that the measurements of a single band are normally distributed with a SD of 2 units (Table 4) then the raw data can be smoothed to produce the distribution (Figure 12). From this graph we can then determine $f(z)dz$ as the probability that a measurement on a band from a person selected at random from the population will lie in the interval $(z, z + dz)$.

probe YNH24
band matching separation (n=111)

%

distance separation (1 unit =0.35mm)
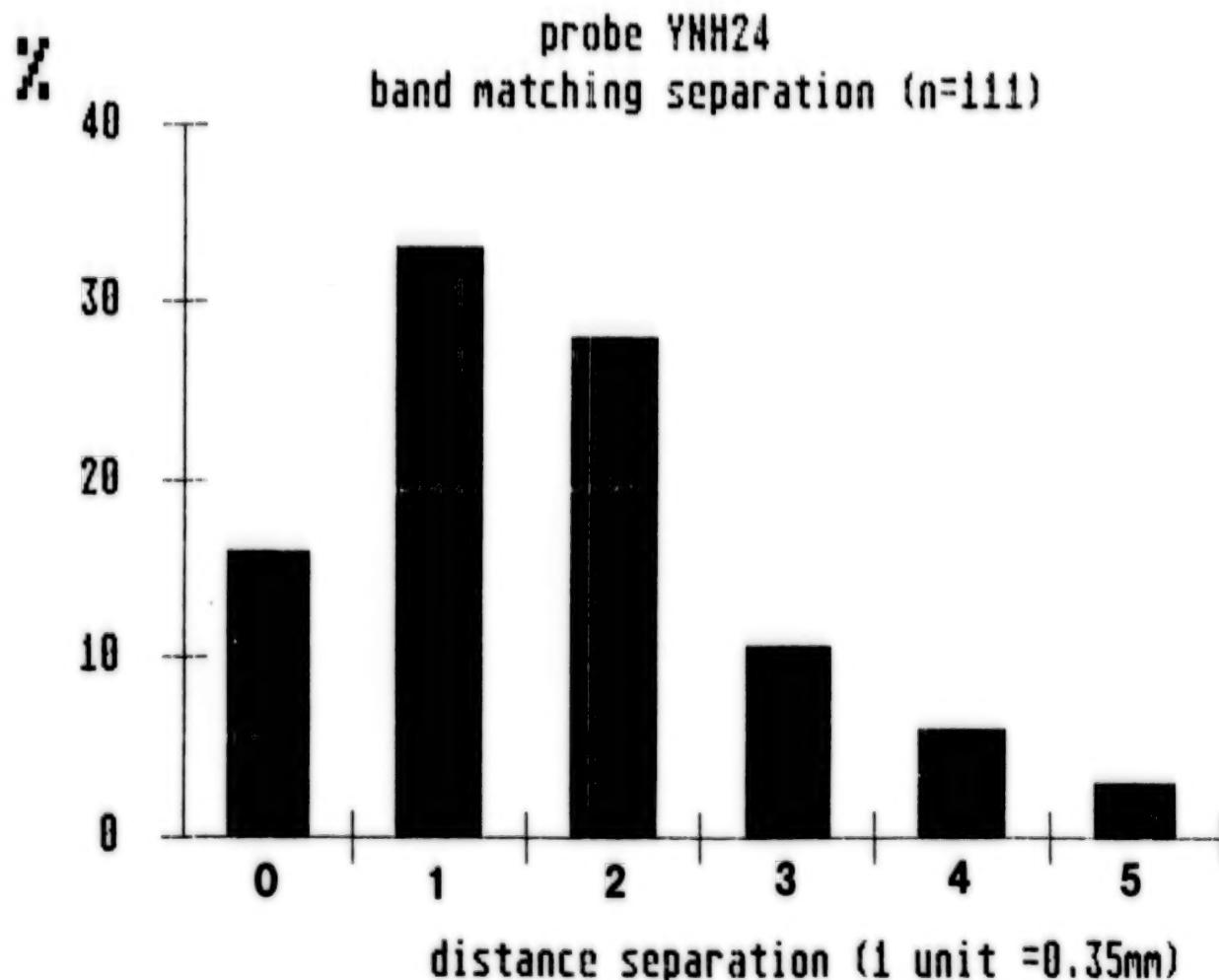
Figure 10. The distribution if differences between pairs of bands of common origin after probing with YNH24. Distance of separation (in position units in this case equivalent to 0.35 mm) is plotted against percentage of band pairs at that separation. As this is blood sample/blood sample comparison the sign of the separation has been ignored and, therefore, the values 1 to 5 comprise the sum of bands separated by +1 and -1 unit, +2 and -2 units, etc. The distribution is not significantly different from a normal distribution.

Consider the comparison of the control blood result with that of, for example, a semen stain from the scene. Single locus analysis is carried out with the result:

$y = (y_1, y_2)$ for the crime stain

$x = (x_1, x_2)$ for the suspect's sample

We test the alternative hypothesis:

C: the crime stain was left by the suspect

C: the crime stain was left by someone else

Following Lindley (1977) we calculate the likelihood ratio:

$$\frac{p(x,y \text{ IC})}{p(x,y \text{ IC})}$$

where the p's denote the relevant probability densities. The likelihood ratio can be simplified approximately to:

$$\frac{p(\sim)}{p(y)}$$

where:

$\sigma = (\sigma_1, \sigma_2)$

and:

$\sigma_1 = (x_1 - y_1)$

$\sigma_2 = (x_2 - y_2)$

$p(\sigma)$ is evaluated using the within individual measurement error distribution and $p(y)$ is evaluated using the population measurement error distribution.

With regard to $p(\sigma)$ we have assumed that measurement error for a given band will have normal distribution with SD∂. When an individual is considered then both bands will be subject to measurement error. However the measurement
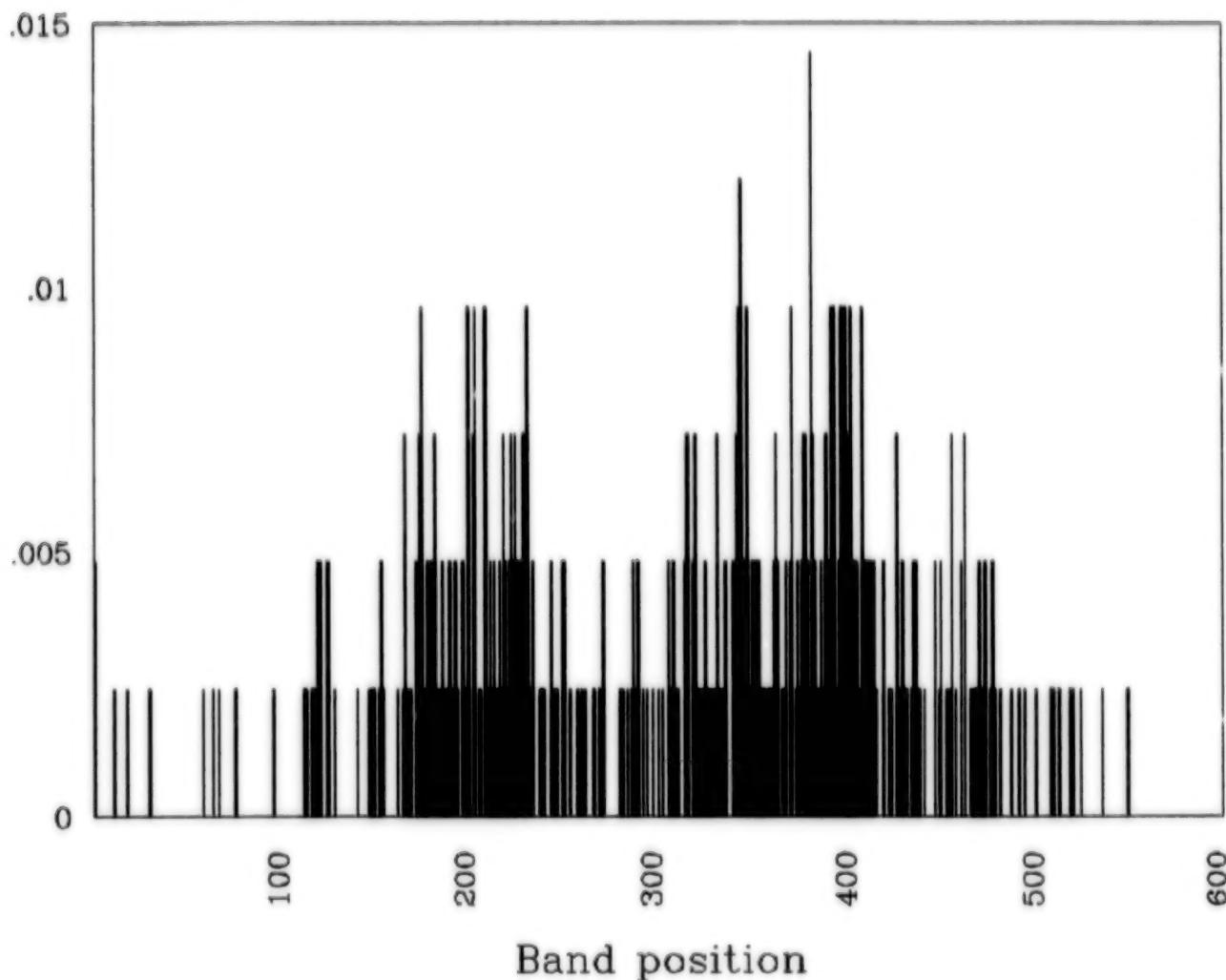
Figure 11. The distribution of bands from 200 individuals. Alu digested genomic DNA was run 30 cm 0.6% agarose gels using Tris-borate buffer. The DNA was then probed with YNH24. Note the gaps in the occurrence of bands, it is presumed that as more samples are analyzed these may be filled. These samples were run as an investigative exercise only, in Europe data bases are likely to be based on HinfI digested genomic DNA.

errors will not be independent of each other: they will be correlated. Thus, for the $p(\sigma)$ the bivariate normal distribution is used.

With regard to $p(y)$ this can be taken directly from the $f(z)$ distribution. The components of $p(y)$, y_1, and y_2, can be sampled independently from the same population provided that a small correction factor is applied to allow for the fact that they cannot be closer than the resolving power of the system. The approximation that results from the within source error components of the two band measurements is considered negligible compared to the overall population variation.

The statistical treatment is system dependent, a laboratory will need to determine the parameters used in the derivations. Also, we recognize that this approach may be sensitive to some of the pa-

rameters involved in the treatment. For example, the bivariate frequency distribution has five parameters and in particular the SD and correlation coefficient may vary with the type of sample analyzed, that is, the SD of blood sample analysis and comparison may be different to that of comparing blood sample and stain analyzed. Furthermore, the distribution of error measurements may not be strictly Gaussian. However, these variables can be modelled and explored using computer simulations and spreadsheets to test the robustness of the treatment. The use of smoothed probability density functions to arrive at estimates of the true probability of band occurrence may at first appear to arrive at strikingly different and generally lower probabilities than the rather crude estimates based upon limited population samples. Further tests and
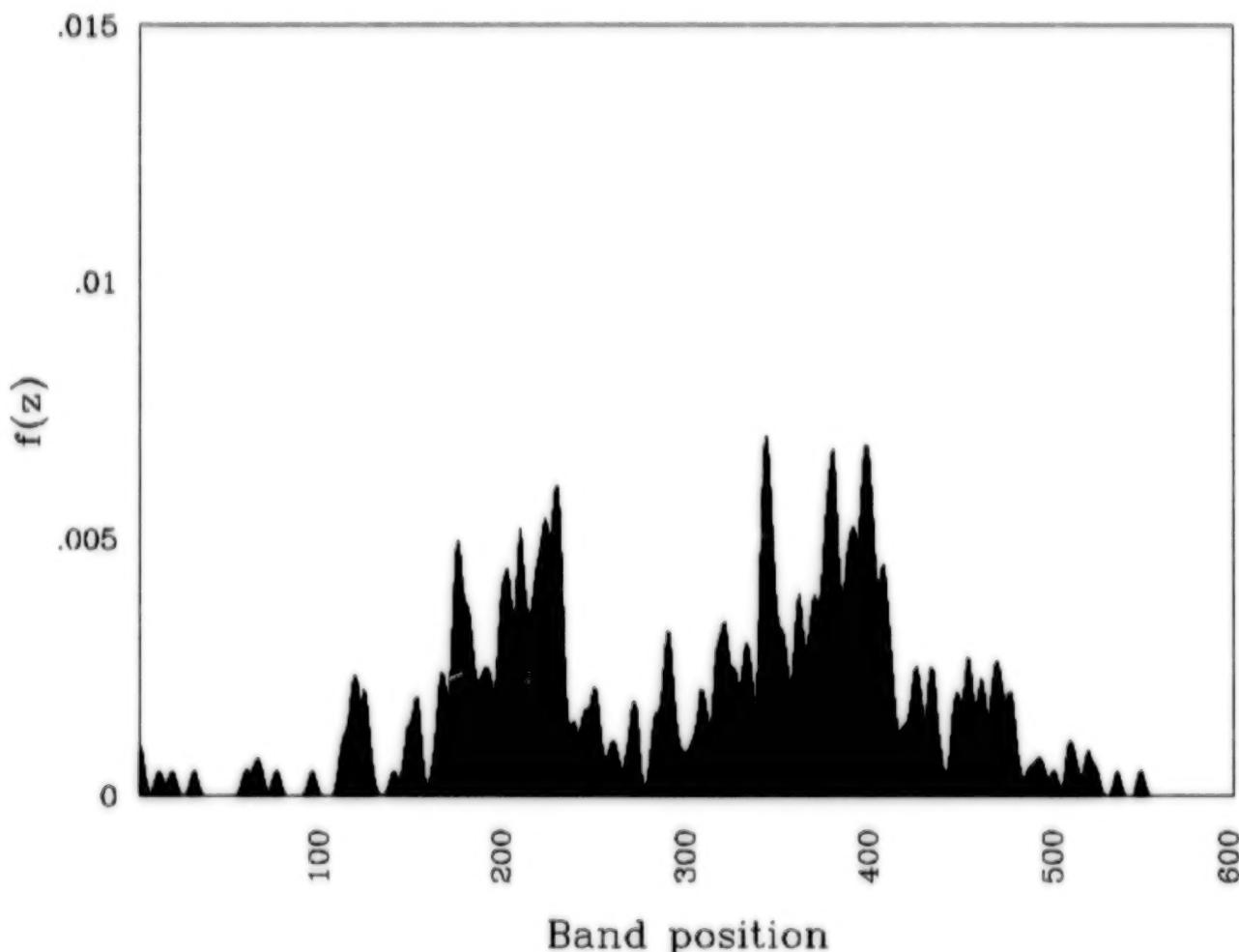
Figure 12. The raw data shown in the previous figure after smoothing using as a smoothing function the measurement error distribution which we have assumed to be normally distributed and in this case has a standard deviation of approx. 2 position units. From the smoothed data we can determine f(z)dz as the probability of a band from a person selected at random from the population will lie in the infinitesimal interval $(z, z + dz)$.

models are required to substantiate the use of the estimates made here, it may also be appropriate to build in conservative safeguards for the use of such models.

Using the likelihood ratio it is possible to assess the quality of the match within the numerator. Furthermore, the use of the correlation coefficient in the bivariate analysis will greatly reduce the value of the numerator, and hence the likelihood ratio, when the two bands in the crime sample are not shifted in the same direction relative to the control sample. This is in concert with the subjective assessment of such potential matches. Thus, the main thrust of this approach is the assistance it offers to the reporting officer in assessing the results. The subjective assessment of a match is removed and is replaced with a treatment that reflects the quality of the results in a likelihood ratio. Currently subjective assessments set the numerator

to 1, it either matches or it does not! As a guide to how it may be expressed in court the following phraseology might be considered:

"The findings have been examined statistically and two possible alternatives has been considered: either the bloodstaining on the clothing has originated from Mr. Smith or the blood has originated from another individual and the observations have occurred by chanced. In this case the findings are at least 14,000 times more likely if the bloodstaining has originated from Mr. Smith than if it had originated from somebody picked at random.

In my opinion the findings therefore strongly support that the blood on the clothing came from Mr. Smith".

In this type of statement the analyst has addressed the meaningful question: What is the probability of the evidence as proposed by Evett (1983). In addressing this question he has considered two

opposing alternatives and in doing so has assigned probabilities to both.

It should be noted that in England and Wales forensic scientists are encouraged to express an opinion as to the strength of the findings hence the final paragraph of the statement.

This paper has dealt with DNA analysis for the immediate future in the HOFSS. Currently under investigation at the CRSE is the use of polymerase chain reaction (PCR) as well as other methods of DNA analysis, it is envisaged that these methods may take their place in operational forensic science in the mid to long term future. However, we envisage that the introduction of such methods, particularly PCR, will require considerable caution.

CONCLUSION

The HOFSS has established MLP DNA analysis in the courts but seeks to introduce SLP analysis using at least four SLPs. It is doing this within a European framework of cooperation: 12 European laboratories are undertaking replicate analysis of *Hinfl* digested genomic DNA using the probe pCRE2.7, derived from YNH24 (Nakamura *et al.* 1987). Data are presented here using AluI digested genomic DNA run on 30 cm, 0.6% agarose gels in Tris-borate buffer. This has been used to investigate the statistical analysis of SLP results and a method, based on a Beyesian type analysis and using density functions, has been proposed. The gathering of a database for statistical analysis of casework is nearing completion, it comprises more than 200 blood samples from each of the three main racial groups: White caucasian, Afro-caribbean and Indian/Pakistani. The samples have been subjected to *Hinfl* digestion and electrophoresis in 25 cm, 0.7% agarose gels run in recirculated Tris-acetate buffer and then probed using pCRE1.2, pCRE 2.7,3'HVR and MS43. The Biotrac machine is being used to analyze the data.

Breakdown of cases that have been processed through the courts, the total 48 is not presented as an exhaustive list but does comprise those cases of which the HOFSS is aware. In most instances forensic scientists present their evidence and then return to their laboratory, thus the information gathered above should be treated as a general, rather than an absolute, guide to the progress of cases in the courts of England and Wales. The cases where the DNA favoured the defence: although this represents a relatively small number of those cases going to court, for cases submitted to forensic science laboratories, where the figure is

known, the suspect is eliminated approximately 30% of the time.

REFERENCES

Baird, M., Balazs, I., Giusti, A., Miyazaki, L., Nicholas, L., Wexler, K., Kanter, E., Glassberg, J., Allen, F., Rubinstein, P. and Sussman, L. (1986). Allele frequency distribution of two highly polymorphic DNA sequences in three ethnic groups and its application to the determination of paternity, Am. J. Hum. Genet. 29:489–501.

Elder, J. K. and Southern, E. M. (1987). Computer aided analysis of one-dimensional restriction fragment gels. In: Nucleic Acid and Protein Sequence Analysis, (Bishop, M. J. and Rawlings, C. J., eds.) pp 165–172.

Evett, I. W. (1983). What is the probability that the blood came from that person? A meaningful question? J. Forens. Sci. Soc. 23:35–39.

Evett, I. W., Cage, P. E. and Aitken C. G. G. (1987). Evaluation of the likelihood ratio for fibre transfer evidence in criminal cases, J. Royal Statistical Soc. (Series C) 36:174–180.

Gjertson, D. W., Mickey, M. R., Hopfield, J., Takenouchi, T. and Terasaki, P. I. (1988). Calculation of probability of paternity using DNA sequences, Am. J. Hum. Genet., 43:860–869.

Gill, P. D., Jeffreys, A. J. and Werrett, D. J. (1985). Forensic application of DNA "fingerprints", Nature 318:577–579.

Gill, P. D. and Werrett, D. J. (1987). Exclusion of a man charged with murder by DNA fingerprinting, Forens. Sci. Intern. 35:145–148.

Gill, P. D., Lygo, J. E., Fowler, S. J. and Werrett, D. J. (1987). An evaluation of DNA fingerprinting for forensic purposes, Electrophoresis 8:38–44.

Hill, W. G. (1986). DNA fingerprint analysis in immigration test cases, Nature 322:290–291.

Jarman, A. P., Nicholls, R. D., Weatherall, D. J., Clegg, J. B. and Higgs, D. R. (1986). Molecular characterization of a hypervariable region downstream of the human alpha-globin gene cluster, The EMBO Journal 5:1857–1863.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985a). Hypervariable minisatellites in human DNA, Nature 314:67–83.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985b). Individual-specific fingerprints of human DNA, Nature 316:76–79.

Lindley, D. V. (1977). A problem in forensic science, Biometrica 64:207–213.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987). Variable number of tandem repeats (VNTR) markers for human gene mapping, Science 235:1616-1622.

Werrett, D. J., Gill, P. D., Lygo, J. E. and Fowler, S. J. (1988). DNA polymorphisms—Practical use, Adv. Forensic Haemogenetics 2:320-338.

Werrett, D. J. and Lygo, J. E. (1989). The role of DNA profiling in the courts, The Law Society's Gazette 13:35-38.

Wong, Z., Wilson, V., Jeffreys, A. J. and Thein, S. L. (1986). Cloning a selected fragment from a human DNA "fingerprint": isolation of an extremely polymorphic minisatellite, Nucleic Acids Res. 14:4605-4616.

Wong, Z., V., Patel, I., Povey, S. and Jeffreys, A. J. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA, Ann. Hum. Genet. 51:269-288.

DNA ANALYSIS IN THE METROPOLITAN POLICE FORENSIC SCIENCE LABORATORY

Brian H. Parkin

Metropolitan Police Forensic Science Laboratory
London, England

Profiling of DNA was introduced into forensic casework in the Metropolitan Police Forensic Science Laboratory in December 1988. Although considerable experience had been obtained with the use of multilocus probes (MLP) 33.15 (Jeffreys *et al.* 1985), Ig51s, a synthetic 400 base pair (bp) probe comprising 25 tandem repeats of the core sequence of 33.15 (Sullivan and Werrett 1988) and 3' α-globin HVR (Higgs *et al.* 1981), a decision was made to use single locus probes (SLP) in casework. Single locus probes appeared to have many advantages over MLPs in forensic casework in that they were more sensitive, they enabled ease of attributing components of a mixture of body fluids to their respective donors and they had the potential to form a computerized index of DNA profiles of criminals and unsolved crimes of violence, particularly rapes.

The SLPs selected for use were MS1, MS31 and MS43A; highly polymorphic probes with heterozygosities varying from 0.96 to 0.99. The probes were purchased from Cellmark Diagnostics, UK. In order to introduce the technology at an early date, Cellmark Diagnostics made available their frequency data bases of Caucasians, Afro-Caribbean and Asian populations (Smith, J. C., personal communication).

Although the numbers of individuals comprising these data bases were low, 200 or fewer, they were used to assist the courts in the evidential value of the results. Eventually population statistics will be built up from unrelated criminals and victims of London's cosmopolitan population, to provide a far larger and more meaningful data base. In order to use this data base, the DNA was restricted with *Hinf*I.

Staff

Casework was initially carried out with a staff of four, two of which were also reporting officers. Due to the heavy caseload, the staff is being increased to seven over the next few months. An extensive training program involving both internal and external training is progressing to enable all the biology reporting officers to report DNA results in their own cases. This training has, of course, put further pressure on the section's resources.

Cases

Cases for DNA profiling have been submitted at a rate of about 40 per month. Initially no selection was made in the type of cases accepted. The policy was to attempt DNA profiling in all cases to rapidly increase the experience in dealing with a range of forensic material. However, to prevent backlogs occurring, the following cases have now been omitted:

(i) Burglaries, where the suspect bled.

(ii) Domestic murders, where conventional grouping could distinguish bloods of the family members involved.

(iii) Brawls, particularly those occurring after heavy drinking.

In addition, requests for DNA profiling in rape cases where no suspect has yet been apprehended are given low priority.

Thus murders and sexual assaults remain as high priority cases.

Samples

In cases of sexual assault, high vaginal and low vaginal swabs were typically submitted. The extracted material from these was combined to obtain the DNA profile.

In violent crimes, bloodstained clothing and weapons, usually knives, were submitted and occasionally cigarette ends and masks for DNA profiling from saliva stains.

Control samples were usually bloods containing EDTA as anticoagulant, although excellent results have been obtained from clotted bloods. Hair and saliva samples have also been used.

Under English law, samples are defined as intimate when they are taken from a body orifice or by means of a hyperdermic needle. An example of a non-intimate sample would be head hair. In order to take intimate control samples, the consent of the suspect and the authority of a senior police officer must be given. Refusal is taken into account by the court. If refusal is given for non-intimate samples in

serious arrestable offenses, a senior police officer may authorize such samples (for example, head hair) to be taken by a police officer who may have to use force. In our experience, control samples from the suspect are normally given.

Method

Samples were extracted with SDS, DTT and Proteinase K with semen DNA being extracted preferentially by the method of Gill *et al.* (1985). The DNA was purified by phenol/chloroform extractions and precipitated by ethanol. In order to further purify the DNA, it was redissolved in 0.2 M sodium acetate and again precipitated with ethanol. The purified DNA was assayed by fluorimetry and after suitable aliquoting, restricted with 30 fold excess *Hinf*I. DNA (0.5 μg) was loaded in to 0.7% agarose gel and after electrophoresis was transferred to a nylon membrane (Amersham Hybond N) by vacuum blotting. The nylon membranes were prehybridized in 150 ml of 0.25 M sodium phosphate buffer, pH 7.2 containing 7% SDS, 1 mM EDTA and 1% BSA for about 10 minutes at 65°C. Hybridization took place at 65°C for 16 hours using the prehybridization solution containing the probe. This solution was labelled with P[32] (dCTP) by the random priming method (Feinburg and Vogelstein 1983).

After probing, the membranes were washed with 2 x SSC containing 0.1% SDS at 65°C followed by washing with 0.1 x SSC containing 0.1% SDS at 65.

In order to calibrate the size of the alleles in the test samples, a series of markers labelled with S[35] (Amersham) were included on each electrophoresis gel.

To conform to quality control standards, all control blood samples were run in conjunction with a bloodstain made from the sample by a person other than the operator. Also a known semen sample was included on each gel.

Autoradiography was carried out using Amersham MP hyperfilm for 1 to 7 days using intensifying screens at −70°C.

Interpretation

When interpreting a set of results, the profile is formed from all three probes, and conclusions were based on this total result. An opinion of match was never made on just one result as there could be a chance of similar size alleles occurring. Occasionally when the preferential semen extraction was not executed, the sample result was compared with the vaginal DNA alleles and the re-

maining alleles considered to be of seminal origin. These were compared with the suspect profile (Figures 1 and 2).
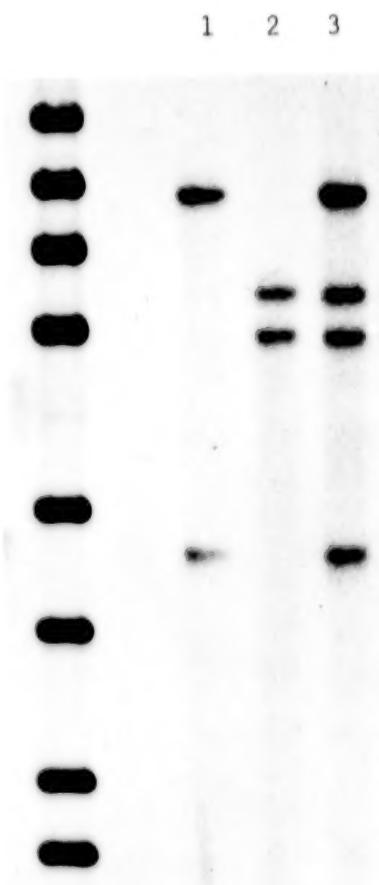
Figure 1. The DNA profile obtained from a semen stained vaginal swab (lane 3) is shown to be composed partly of DNA from the female victim (lane 2). The remainder of the profile matches the DNA control of the suspect (lane 1).(0). 5 μg of DNA loaded per lane. Autoradiography for 6 days. Probe: MS 31).

Determination of allele size is, at present, carried out manually by comparison with the DNA size markers. Using the high molecular weight alleles as a guide, a base line is constructed and measurement of the distance travelled by the marker and sample DNA bands is made. The kilobase size of the sample alleles are then calculated by the method of Elder and Southern (1987).

Results

Results are given for the first 106 cases examined by this laboratory. Profiling of DNA was executed on body fluid stains on a range of substrates composed of natural and synthetic fiber, metal, wood, paper, etc., up to 4 years old. The stains were mainly of semen and blood, but other tissues were occasionally profiled, particularly hair,
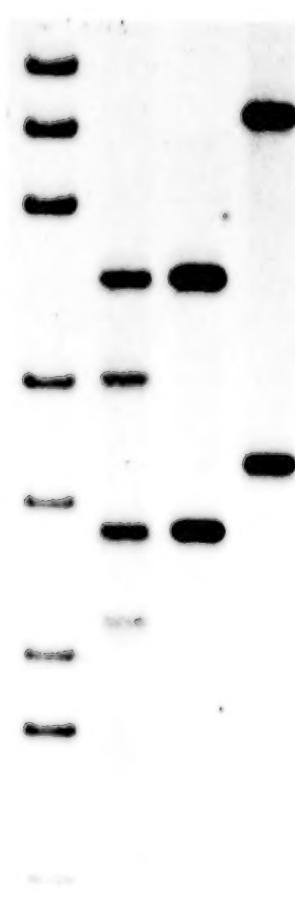
Figure 2. The DNA profile obtained from a semen stained vaginal swab (lane 1) is shown to be composed partly of DNA from the female victim (lane 2). The remainder of the profile does not match the DNA control of the suspect (lane 3) showing he is not the donor of the semen. (0.5 μg of DNA loaded per lan. Autoradiography for 7 days. Probe : MS1)

muscle and saliva. Menstrual blood proved to be a rich source of DNA, hence results were obtained from apparently weak stains.

A readable profile was obtained in about 70% of cases (Table 1), although this could have been improved if selection of samples had been carried out. A breakdown of this figure shows that approximately 77% of cases involving semen were successfully profiled but only about 66% of cases where the test material was bloodstaining. Although a higher rate of success was obtained with semen stains, it was encouraging to see DNA profiles obtained from a higher number of forensic bloodstains than previously thought.

Table 1. BREAKDOWN OF 106 CASES IN WHICH DNA PROFILING WAS CARRIED OUT.

Cases with semen stains	64	
Those with readable semen DNA profile	49	(76.6%)
Cases with blood stains	41	
Those with readable blood DNA profile	27	(65.8%)
Cases with hair	5	
Cases with muscle	8	
Cases with menstrual blood stains	2	
Cases with saliva	5	
Cases with readable DNA profile	74	(69.8%)
Cases with no DNA profile	32	(30.2%)

An analysis of the control bloods and laboratory made stains shows that a higher success rate was achieved when profiling the laboratory made stains (94.9%) than for the bloods from which they were made (91.5%). No reasonable explanation has been advanced to date.

The frequency of these alleles in the population being considered are calculated by references to the data tables. To avoid underestimating the frequency of any allele, a distance of 1mm each side of the estimated allele size is taken into account. This 2 mm error window covers a spread of DNA fragment sizes varying from 100 bp for alleles falling within this error window. The fragment sizes are then added and taken as the final frequency. In addition, when the frequency of any allele falls below the mean allele frequency (Wong et al. 1987), the result defaults to this mean frequency. Using this approach, the following range of results have been obtained (Table 2).

Table 2. RANGE OF PROBABILITIES OBTAINED FROM 60 CASES IN WHICH THE DNA PROFILE FROM THE SUSPECT MATCHED THAT OF A CASE STAIN.

Probability	No. of Cases
1 in 1 to 500,000	12
1 in 500,000 to 1 million	6
1 in 1 to 5 million	16
1 in 5 to 10 million	9
Above 1 in 10 million	17

Probes used: MS1, MS31 and MS43A.
Results are expressed as the probability of finding a second person with a DNA profile matching that of the suspect.

Semen Stains

Although the number of spermatozoa present in the stain is the most important factor in determining the likelihood of obtaining a DNA Profile,

a breakdown of time after intercourse was also investigated (Table 3).

Table 3. BREAKDOWN OF RESULTS FROM SEMEN STAINED VAGINAL SWABS.

Concentration of spermatozoa	1+	2+	3+	4+
Results	*p. n.	p. n.	p. n.	p. n.
Time after intercourse				
0-2 hours	1 2	- -	- -	4 -
2-5 hours	2 1	5 -	5 1	2 -
5-10 hours	1 -	2 1	2 1	
10-15 hours	- 1	4 -	2 1	
15-24 hours	1 -	1 -	1 -	

(*Where p = positive result and n = negative result).

The concentration of spermatozoa were scored as described by Davies and Wilson (1974) on a sliding scale where 4+ showed a high concentration and 1+, a low concentration.

Results showed that vaginal swabs were normally taken from the victim within 24 hours of the offense, and that DNA profiles were normally obtained for concentrations of 2+ and above. The success rate for semen stains with concentrations of 1+ was less successful.

On very few occasions were the examination of oral or anal swabs requested. Of these, an occasional successful profile was obtained. Saliva stains were also few in number and although, in general, insufficient DNA was extracted, a readable profile was obtained from a cigarette end in one instance (Figure 3). A successful profile was also obtained from nasal mucus.

Future

At present, the technique and efficient management of DNA profiling at this laboratory is being consolidated, coupled with a reorganization of staff thus allowing for a contraction of conventional immunological and biochemical systems for body fluid grouping. However, a core set of groups will continue to be used to give a rapid result when required to screen in cases where only two persons could have bled (for example, domestic disputes) and to group stains of insufficient size for DNA profiling.

Certain problems have been highlighted after 6 months experience and these will form the basis of future research:

1. Improve turn around time.

Figure 3. The DNA profile obtained from saliva on a cigarette end (lane 3) matches profile of suspect (lane 1). Lane 2 shows part of DNA size markers. (0) 5 μg of DNA loaded per lane. Autoradiography for 8 days. Probe: MS 31).

2. Automate labor intensive stages.

3. Investigate more sensitive and informative probes.

4. Establish large frequency data bases for a number of ethnic populations.

5. Begin to develop data bases of profiles from criminals and unsolved crimes.

European Integration of DNA Profiling

In 1992 it will be far easier to travel between countries forming the European Economic Community (EEC) when border restrictions are relaxed. This means that criminals can also move freely throughout Europe. At the instigation of Dr. Brian Sheard, Director of this laboratory, a meeting of representative European Laboratories known to be working on forensic DNA profiling took place in London in October, 1988. Laboratories from Denmark, Germany, Italy, Netherlands, Switzerland and U.K. attended this first meeting, and it was unanimously agreed that standardization was necessary before laboratories took irreversible decisions.

Further, to obtain minimum standards, research efforts should be shared, quality assurance and quality control adhered to and where possible, standard methods used so realistic comparisons can be made.

A small working group was set up to consider the important facets of a common restriction enzyme and common probes. After hearing their recommendations at the second meeting, hosted by Professor Brinkmann in Munster, it was agreed to use HinfI as a restriction enzyme and YNH 24 as a common probe.

In the next meeting, at the Hague, it is hoped that further steps in standardization will be achieved, including the identification of a core package of probes. It is also hoped that more countries will join the group.

REFERENCES

Davies, A. and Wilson, E. (1974). The persistence of seminal constituents in the human vagina, Forens. Sci. 3:45–55.

Elder, J. K. and Southern, E. M. (1987). Nucleic Acid and protein sequence analysis—a practical approach, (Bishop, M.J. and Rawlings, C.J., eds.), IRL Press, Oxford. pp. 165–172.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity, Anal. Biochem. 132:6–13.

Gill, P., Jeffreys, A. J. and Werrett, D. J. (1985). Forensic application of DNA "fingerprints", Nature 318:577–579.

Higgs, D. R., Goodbourn, S. E. Y., Wainscoat, J. S., Clegg, J. B. and Weatherall, D. J. (1981). Highly variable regions flank the human alphaglobin genes, Nucleic Acids Res. 9:4213–4214.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985). Individual specific "fingerprints" of human DNA, Nature 316:76–79.

Sullivan, K. and Werrett, D. J. (1988). Determination of the sequence of a DNA probe "Ig51s" by the Sanger-dideoxy chain termination technique, CRE Technical Note 626 No. 61900.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA, Ann. Hum. Genet. 51:269–288.

BLANK PAGE

THE USE OF DNA POLYMORPHISMS IN THE POLICE LABORATORIES OF THE FEDERAL REPUBLIC OF GERMANY

Hermann Schmitter

Bundeskriminalamt
Wiesbaden, Federal Republic of Germany

Sigrid Herrmann

Polizeitechnische Untersuchungsstelle,
Berlin, Federal Republic of Germany

Werner Pflug

Landeskriminalamt
Stuttgart, Federal Republic of Germany

Among forensic scientists working on the field of examination of bloodstains and stains of body fluids other than blood there is no doubt on the high potential of DNA polymorphisms that can be used in forensic practice. Therefore, it was found to be necessary to introduce the new technology even in the forensic science laboratories of the police forces of the Federal Republic of Germany and of West Berlin.

For your understanding it may be useful to give a brief description of the institutions and laboratories working on the field of forensic examinations of bloodstains and stains of body fluids other than blood in our country.

According to the organizations to which they belong there are different kinds of institutes:

—The institutes of legal medicine are departments of many universities of Germany. There are about thirty laboratories. In the field of forensic serology these are primarily involved in paternity testing but many of them also analyze stains. All of these institutes are independent on each other but come together for annual meetings.

—Only few private laboratories are involved mainly doing paternity testings.

—Ten laboratories in Germany are run by police organizations.

The Federal Republic of Germany consists of ten states plus the city of West Berlin. Each state has its own police organization including the branches of crime investigations. All of them are independent of each other.

In addition to the state police organizations, there is a federal police for crime investigation called the Bundeskriminalamt (BKA).

Nine of the states, West Berlin and the BKA have set up forensic science institutes each of which analyzes physiological stain evidence in cases of crime.

Although the police laboratories are independent of each other the collaboration among the people working in the police laboratories is excellent. Symposia and workshops are held periodically on all areas of forensic sciences.

In this paper we will concentrate on the steps performed by the police laboratories to introduce DNA analysis into their routine casework.

The heads of the forensic science institutes of the states and of the BKA attended a conference in 1986 and established a DNA analysis program. Three scientists, the authors, due to their education and familiarity with the principles of molecular biology were appointed to do the basic work as necessary for the introduction of this new methodology.

The duties of the group included the following items:

1. To obtain experiences on the different steps of preparation in the analytical procedure;

2. To choose suitable systems of restriction enzymes and probes applicable to forensic casework;

3. Following the assessment of the method to train the colleagues of the remaining state laboratories.

Our group did not receive orders to perform research work on the applications of DNA techniques. Rather, we were authorized to contact laboratories that already had established DNA profiling systems.

We started working using standard protocols of conventional methods for DNA extraction, restriction, electrophoresis, Southern blotting as well as labelling and hybridization. That work was done

by each member of the working group in their home laboratory to get familiar with the techniques. Based on this preliminary work a one week workshop was conducted in the laboratory of Berlin in November, 1987. Participants in that workshop were the authors and Mrs. Marie Luise Sonntag a technician of medicine working in the BKA.

At that time just two different probes were available to us:

—The probe pHMF provided to Dr. Hermann by professor Hoar of the University of Toronto.

—The probe 3'alpha hemoglobin provided by Dr. Dave Werrett of the CRSE in Aldermaston.

During that workshop we only obtained reasonable results using the pHMF probe. The results indicated that using *Hinf*I as the restriction enzyme, the probe probably was not very useful due to an obviously high band sharing rate as shown within four randomly chosen unrelated individuals. Nevertheless, the results obtained at that workshop encouraged us to get in contact with colleagues who already were working on DNA profiling systems.

Thus, since March 1988, our group has visited the laboratories of ICI in Gadbrook Park, of Cellmark in Abingdon and of the CRSE in Aldermaston.

These visits have been very informative for us as well as very helpful for our further work. Following our visit to ICI, we were provided with the multi locus proves 33.15 and 33.6 as well as with the single locus probes MS1, MS31, MS43 and g3, from that organization. Since then nearly all further experiments performed by our group were done using *Hinf*I as the restriction enzyme and the probes previously mentioned.

At Aldermaston where we were allowed to stay for one week, our colleagues gave us the opportunity to practice the whole procedure as it was done at that time in the laboratory of the CRSE. That training has given to us a realistic impression of how DNA profiling techniques can be used on routine casework.

Following our visits to the laboratories in England we performed further experiments at an advanced level. At the end of 1988 we met again to run another workshop in Berlin. The number of participants at that workshop as well as the time, were extended compared to the first one showing the increase of acceptance of DNA analysis within the forensic science institutes of West Germany. On that second workshop that took three weeks Dr. Seiffert who joined the laboratory of Berlin

and the technicians Mrs. Eberspacher of Baden-Wuerttemberg and Mrs. Schiel of Berlin were added to the participants of the first workshop.

The results obtained in that workshop demonstrated that we generally were able to apply DNA analysis to casework. However, we decided to begin with the application of RFLP techniques only in cases, where after having done conventional analysis, sufficient material has been left for DNA analysis. The reason for the limitations, according to the kind of suitable cases was that unless satisfactory results could be obtained from the second workshop it could be very dangerous to apply new techniques without having acquired experience in actual casework. This has been our strategy in the past, prior to adding a further marker system to our routine testing program.

A report of the results of the workshop and of the conclusions drawn was prepared for the heads of the institutes. They agreed to our conclusions and decided to give us additional time necessary for getting those experiences. At this point it has to be mentioned that in the meantime many cases sent to us to be analyzed using the new technique were passed on to Cellmark Diagnostics by us after we had found the material to be suitable for that kind of analysis.

Thus, to gain experience in actual case work we have analyzed stain materials from a number of sexual assault cases, that had been completely finished. In addition, we have started a proficiency testing program where the three laboratories involved have to analyze dried blood samples from thirty different individuals to see whether they will come up with the same results. Thus, now we feel we are ready to apply the technique in actual casework.

In parallel the training of the colleagues of the remaining state laboratories has been started.

This description provides a history of the manner in which DNA analysis has been established in the police laboratories of West Germany. Hopefully, this discussion provides some useful information for other organizations planning to introduce DNA analysis.

As previously explained it was not our intent to perform any research work on DNA analysis. We almost followed the standard procedure mainly those described by Dave Werrett in the manual as it is used in Aldermaston. Nevertheless, trying to increase our experience using the new technology we found it necessary to do some basic studies resulting in some minor but interesting findings.

However, we will first provide a listing of the standard procedure used in our laboratories:

1. Unless preliminary extraction of mixed stains has to be done for the extraction of DNA from stains, the materials carrying the dried substances are immersed in Eppendorf tubes each in 0.5 ml of lysis buffer (that is, high TEN containing DTT, SDS and proteinase K). The mixtures are allowed to incubate over night at 37°C.

2. The residues are removed by extraction using standard phenol/chloroform mixtures. The extractions normally are repeated once.

3. After addition of a tenth of volume of 2 M Na-acetate to the aqueous supernatants the DNA's are precipitated by adding two fold volumes of ethanol. The tubes are left at -20°C in a freezer for three to four hours. The tubes then are centrifuged and the pellets are washed with 70% ethanol and then dried either by leaving them on the benchtop over night or by evaporating under vacuum for about 10 minutes.

4. After drying the pellets are redissolved. We normally add 0.02 ml of distilled water to each extract. The solutions then are dialyzed for 2 hours against dialysis buffer, that is, 0.1 M Tris-HCl, pH 7.2 using millipore filters as dialysis membranes.

5. Following dialysis the volumes of the dialysates are measured thoroughly. All dialysates are then adjusted to an appropriate volume and aliquots are removed to run a control gel.

6. To the remaining solutions suitable volumes of digestion buffer, of spermidine and of restriction enzyme are added and the mixtures are incubated over night at 37°C. If the volumes are in a reasonable range (that means up to 0.07 ml) 0.01 ml of stop mix containing bromephenolblue are added and the mixtures immediately are applied to the gels for electrophoresis. If the volumes are higher due to the uptake of water during dialysis the restricted DNA again will be precipitated and redissolved in an appropriate volume of distilled water before using the extracts for electrophoresis.

7. The DNA fragments are electrophoresed over night in 0.7% or 0.8% agarose gels. The running distance is monitored next morning. When necessary the run will be continued until the desired conditions are met for example until the 2.0 kilobase band of the size marker used has reached the anodal end of the gel.

8. Following depurinization in 0.25% HCl denaturation in 0.4 M NaOH/ 3 M NaCl and neutralization in 1.25 M tris-HCl the DNA is transferred to a nylon membrane by Southern blotting. The nylon filters used so far in our laboratories were Gene Screen Plus, Zeta Probe and Nytrane.

9. For the labelling procedures of the probes as well as for the conditions of hybridization we follow exactly the protocols given by the manufacturer of the probes used, that is, in most of our experiments done so far we have applied the recipes given to us by ICI.

We will now describe some of the findings obtained in our experiments. It appears to be obvious that the most important step in the analysis is the extraction procedure, for the quality of the results that can be obtained mainly depends on the amount as well as on the quality of extracted DNA. One prerequisite for an efficient extraction is the balanced relationship between the amount of material to be extracted and the volume of extraction buffer. To get an estimation of the optimum relationship we have cut defined circular pieces out of areas of sheets of cotton cloth homogeneously stained with blood using a punch. In equal volumes of 0.5 ml each we have extracted, DNA was estimated by comparing the fluorescence of ethidium bromide following electrophoresis.

These experiments showed that in the volume used, the yield of DNA extracted increased up to the sample containing 9 discs, that is, about 3 cm² of cotton cloth. Above that amount no further increase was observed. When an excess of discs, up to 15, was added, the yield of DNA decreased significantly.

To do the first extraction with phenol/chloroform mixture we do not remove the carrier of the stain for it normally remains in the phenolic residue. After having removed the aqueous supernatant we extract the phenolic phase for a second time using extraction buffer, that is, high TEN. We have found that in most cases this second extraction yields about the same amount of high molecular weight DNA as is obtained by the first extraction. However, the aqueous supernatant of a third extraction did not contain reasonable amounts of DNA. Thus, the efficiency of extraction significantly can be increased by reextraction of the residue of the first wash with phenol/chloroform. The DNA obtained of the second extraction was found to exceed the same properties according to molecular size as well as to digestibility as the DNA of the first extract. After precipitation and redissolving in distilled water the DNA yields of both extracts can be combined and treated as one extract through the further steps of the analysis.

In order to get information on both the amount and the purity of the DNA extracted from

stain material we have taken UV spectra of our DNA extracts. The measurements were done with a microscope spectrophotometer as used in our textile fiber sections. To be able to pick up the spectra of the solutions we have constructed microcuvettes made of quartz glass wherein 0.02 ml of a solution could be applied. As previously mentioned the extracted DNA from stains is redissolved after precipitation in 0.02 ml of distilled water. Thus, with the microcuvettes we are able to measure the whole extract. Following spectrophotometry the extract can be removed from the cuvette and transferred back into the tube resulting in a minimum loss of material. In nearly all extracts from stains we have found contaminants exceeding high absorbances in the range of short waved UV light with a maximum at about 240 mn. We neither had the time nor the opportunity to analyze the contaminants that seemed to inhibit the restriction enzyme *Hinf*1. On the other hand the substance could be removed by dialysis indicating they were low molecular weight. After dialysis the digestion problems disappear.

Using different membranes for the Southern transfer as just mentioned we have not observed significant differences with respect to their properties as used in hybridization experiments with probes labelled with radioactive nucleotides.

On the other hand the three membranes used so far in our laboratories exceeded great differences using nonradioactive labelled probes. In experiments we have done using the nonradioactive labelling kit of Boehringer, Mannheim we have found Zeta Probe to be unusable due to a homogeneously dyed dark background. Gene Screen Plus in the same series of experiments led to clear results with good contrasts even if the back side of the membrane was dyed by the staining mixture. Of the three membranes tested, the best results were obtained with Nytran filters. On these membranes clearly contrasted bands could be observed with negligible background staining.

In conclusion we would like to make some remarks on single locus probes (SLP) and multi locus probes (MLP).

When working on DNA polymorphisms always keeping in mind that the analysis must be applicable in forensic casework. We have found single locus systems to be preferable to multilocus probes for different reasons:

1. Single locus systems (SLS's) are more sensitive than MLP's. Many of the blots have given good results after we have hybridized them with SLP's. On the same blots we have seen primarily poor results after reprobing with MLP's. Thus, in casework more stains will be suitable for DNA analysis using SLS's.

2. Some of the artifacts occur for different reasons. For example, partial digestion, and mixing of samples during the purification procedure. Contamination of semen stains on vaginal swabs can easily be detected using single locus probes.

3. The probability of exclusion of a falsely suspected person using just one SLP is very high with respect to the probes mentioned from ICI. Thus for exclusion one probe is sufficient with a very high probability.

4. In case of inclusion the value of matching using one SLP is sufficient to confirm preliminary suspicions. The values are as high as they can be obtained by conventional analysis within the same time. The individualization capacity as stated for the so called DNA fingerprinting using MLP's can be achieved by reprobing with three or four additional SLP's.

Thus, we have decided to concentrate our further work on the use of SLP's.

In conclusion we find that the application of RFLP techniques in forensic casework appears to be a great success for the evaluation of physiological stain evidence. The methods used require fundamental knowledge of the biochemistry of nucleic acids as well as of molecular and general genetics. Thus, an appropriate amount of time for education and training as well as for work to collect experience is necessary before the methods can be applied cases. Otherwise the benefits of the new technique can prove to be one of the greatest blunders in the history of the so called Forensic Serology.

DNA ANALYSIS IN THE FBI LABORATORY

Dwight E. Adams, Lawrence A. Presley, Harold A. Deadman and Audrey G. Lynch

FBI Laboratory
Washington, D.C.

The application of DNA analysis using restriction fragment length polymorphism (RFLP) for the examination of forensic evidence began in the DNA Analysis Unit (DNAU) at the FBI Laboratory in December, 1988. Implementation of this technology was the culmination of approximately three years of research by scientists at the Forensic Science Research and Training Center (FSRTC), FBI Academy, Quantico, Virginia. Researchers at the FSRTC worked in concert with numerous individuals from many disciplines of science to bring to fruition DNA analysis in criminal cases.

The RFLP method developed by the FBI has been described in the literature (Baechtel 1988; Budowle 1988; Budowle *et al.* 1988). Of importance to forensic science prior to the introduction of a novel technology is the assurance that the DNA profile results obtained are both reliable and reproducible. Considerable effort was first extended in the validation of the DNA typing method which addressed such issues as biological and chemical contamination, environmental insults and mixed body fluid stains (Adams 1988). Secondly, an aggressive quality control (QC) program was adopted within the DNAU to monitor the entire RFLP system (Table 1) and the interpretation of results (Monson 1988; Mudd and Presley 1988).

Table 1. QUALITY CONTROL IN RFLP ANALYSIS IN THE DNA ANALYSIS UNIT, FBI LABORATORY.

YIELD TEST GEL

A. Visual Marker—Lambda *Hind*III
- determines approximate size of extracted DNA
- fragment sizes range from 2000 to 23,000 bp

B. Yield Calibration Set—Lambda DNA
- intensities of bands can be compared to samples to estimate quantity and quality of DNA
- concentration dilutions range from 10 to 300 ng DNA

C. Human DNA Cell Line Control
- monitors extraction efficiency of known cell quantity
- monitors overall RFLP analysis system

Table 1. QUALITY CONTROL IN RFLP ANALYSIS IN THE DNA ANALYSIS UNIT, FBI LABORATORY.—Continued

POST RESTRICTION TEST GEL

A. Visual Marker—Lambda *Hind*III
- same as above

B. Digestion Control—Human DNA cell line
- measures restriction enzyme reliability
- restriction enzyme also tested by lot

ANALYTICAL GEL

A. Combination Marker
- visual marker—adenovirus DNA composed of 9 fragments with a range of 1 to 8 Kb
- determines completeness of electrophoresis
- molecular weight size marker

B. Molecular Weight Size Marker—digests of viral DNA's
- used to determine sample fragment size
- ranges from 600 to 13,000 bp

C. Allelic Control—Human DNA cell line
- provides a control of electrophoresis, sizing process and probe determination
- monitors hybridization efficiency and loss of DNA following stripping and rehybridization
- overall system check—same source of DNA for yield test gel, post restriction and analytical gel

EQUIPMENT

A. Service Maintenance Contracts
- Kodak X-Omat film processor
- Autoclave

B. Quarterly Monitor
- Analytical balances
- Pipets

C. Daily Monitor
- Refrigerator and freezer temperatures
- Water bath temperatures
- Oven temperature
- Environmental shaker temperature
- pH meter—by known standards
- Electrophoretic equipment—use of calibrated markers
- Nanopure water system—monitors water quality

The most important internal QC used in DNA typing is the human cell line. This internal control is treated in the same manner as known and questioned stains and serves to monitor organic extraction performance, digestion with the restriction endonuclease, electrophoretic separation and ultimately the interpretation of the final autoradiog-

raphy results. Other QC's utilized with the DNAU are described in Table 1.

Quality assurance is regularly monitored in the DNAU by two types of proficiency tests. The first is an open proficiency in which samples are prepared quarterly at FSRTC and submitted for analysis. These samples are known to the examiners as proficiency samples. The second type of proficiency test is considered blind. These samples, also prepared by FSRTC, are submitted by a cooperating law enforcement agency and simulate an actual case in every way. The extent to which the blind samples are conducted is unknown to DNAU examiners.

Prior to implementation, the DNAU of the FBI Laboratory adopted a case acceptance policy which limits the examination of evidence to violent crimes with known suspects. Analysis of DNA is performed on evidence submitted by state or local crime laboratories lacking DNA typing capabilities following serological screening. Cases accepted without prior serological analysis are first processed in the Serology Unit of the FBI Laboratory for the identification of semen and/or human blood.

In the first six months of operation, there had been approximately 300 cases accepted for examination by the DNAU. These cases represent submissions from federal (10%) and state (87%) law enforcement agencies as well as United States territories and possessions (3%). Analysis of DNA has been requested by 46 states.

A REVIEW OF THE FIRST 100 CASES SUBMITTED FOR DNA ANALYSIS

Of the first 100 cases received in the DNAU, 71% were rapes, 27% homicides and the remaining 2% consisted of other violent crimes. Serial rapes accounted for 16 of the 71 rape cases. An examination of the RFLP results for the first 100 cases determined that 87 cases yielded successful DNA profile results. We define a successful result to include at least one probe result from a probative questioned body fluid stain to be compared to a known blood sample within a particular case. The remaining 13 cases yielded degraded or insufficient DNA. The variable number tandem repeat sequences recognized by the probes utilized in the DNAU exhibit a degree of polymorphism such that approaching individualization of a sample requires analysis with at least three or four probes (Budowle 1988). The DNAU obtained results from four probings in 64 of the first 100 cases. Our results suggest that a successful initial probing is likely to be followed by additional successful probings. Three successful probings were obtained in 14 cases, 2 probings in 8 cases and only 1 case yielded a single probe result.

One of the most significant contributions of DNA analysis to law enforcement is the ability to distinguish contributors of a mixed body fluid stain when compared to known blood samples. The differential DNA extraction procedure adopted by the FBI (Baechtel 1988) allows for the separation of the male and female DNA components in mixed body fluid stains such as semen found on vaginal swabs. Of the first 71 rape cases submitted for DNA analysis, DNA profiles from 76% matched the suspects while 24% showed conclusively that the suspect was not the contributor of the semen. In all of these cases, limited serological analysis did not eliminate the suspect as a possible source of the semen due to the inherent difficulties in antigen detection from body fluid mixtures.

CASE SYNOPSES

The following represent synopses of seven different case scenarios encountered in the DNAU during the first six months of operation. Specific details from some examples have been purposefully omitted while others contain hypothetical information to protect the integrity of cases still under investigation and/or litigation.

HOMICIDE

In a Southeastern United States community, two young girls were abducted, sexually assaulted and one was murdered. A suspect was later identified and a search conducted at his residence located a blanket on which was identified human blood. The DNA analysis performed by the FBI Laboratory determined that the DNA profiles obtained from these bloodstains matched the DNA profiles from the murdered victim (Figure 1).

RAPE (SUSPECT MATCHES)

A Louisiana town was the location of a rape and robbery of a young female. A suspect was developed through informant information and the victim and witness later identified him through a photographic line-up. However, they were unable to identify the suspect in a physical line-up. Serological testing conducted by the state crime laboratory identified semen consistent with the suspect through antigen and protein markers. DNA analysis was performed by the FBI Laboratory on the vaginal swabs and compared to known blood sam-
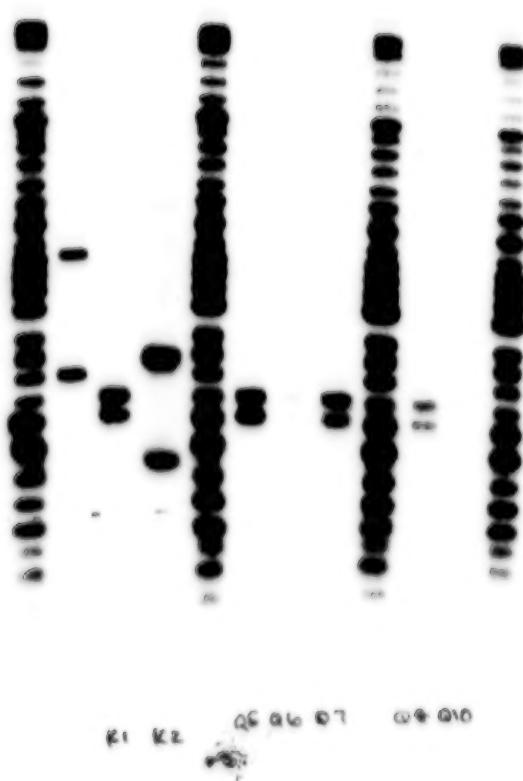
Figure 1. Autoradiogram results from hybridization with single locus probe D2S44. Lanes 1, 5, 9 and 12 contain the molecular weight size markers. Lane 2 contains the human cell line control. Lanes 3 and 4 (K1 and K2) and the known blood samples from the victim and suspect, respectively. Lanes 6, 7, 8, 10 and 11 (Q5, Q6, Q7, Q8, and Q11) are the questioned bloodstains.

Figure 2. Autoradiogram results from hybridization with single locus probe D17S79. Lanes 1, 5, 9 and 12 contain the molecular weight size markers. Lane 2 contains the human cell line control. Lanes 3 and 4 (K1 and K2) contain the known blood samples from the victim and suspect, respectively. Lanes 6 and 7 represent the female fraction DNA of two vaginal swabs (Q1 and Q2) and lanes 10 and 11 are the male fraction DNA from the same two vaginal swabs.

ples of the victim and suspect. The differential extraction procedure was utilized to separate the male and female DNA found in the mixed body fluid stain. The DNA profiles produced from the semen matched the DNA profiles from a known blood sample of the suspect (Figure 2).

RAPE (SERIAL)

In a series of rapes in Illinois, the DNA profiles from semen found on the victim's clothing matched the DNA profiles from the known blood sample from the suspect. Even without a suspect identified, it would have been possible to determine that the same individual was responsible for each rape. In one of the cases, the victim never saw the face of the suspect, so a visual or photographic identification of the suspect would not have been possible. In each case the suspect would grab the victims from behind and drag them to a secluded area. He would then proceed to rape, rob and beat the victims. The suspect lived within a short dis-

tance of the area in which the rapes occurred and he had a previous record of sexual assaults.

RAPE (MULTIPLE SUSPECTS)

In December, 1985, the body of a female was found in a California canal. She had been raped, murdered and found bound with duct tape. An autopsy was performed and a rape kit was collected. Conventional serological testing of the vaginal swab indicated the presence of semen. The ABH blood group substances were consistent with the victim and no PGM activity was detected. Recently, two male suspects were identified. Both suspects were serving prison terms in California and Georgia when identified as possible suspects. The vaginal swab, which had been frozen since 1985, and the known blood samples from the suspects were sent to the FBI Laboratory for DNA analysis. Both suspects were identified as contributors of semen from the vaginal swab (Figure 3).
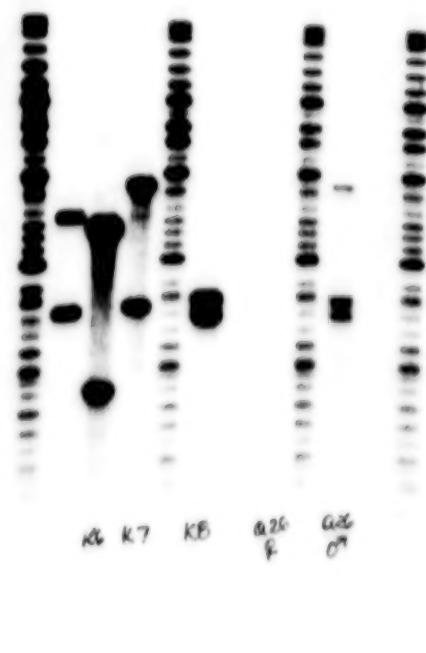
Figure 3. Autoradiogram results from hybridization with single locus probe D2S44. Lanes 1, 5, 9 and 12 contain the molecular weight size markers. Lane 2 contains the human cell line control. Lane 3 (K6) represents the known blood sample from the victim and lanes 4 and 6 (K7 and K8) are the known blood samples from the two suspects. Lane 8 is blank. Lane 9 (no results) represents the female fraction DNA from a vaginal swab and lane 10 is the male fraction DNA from the vaginal swab. Note the four bands in lane 10 representing a mixture of DNA from the two suspects.

Figure 4. Autoradiogram results from hybridization with single locus probe D17S79. Lanes 1, 5, 9 and 12 represent the molecular weight size markers. Lane 2 is the human cell line control. Lanes 6, 7 and 8 are the known blood samples from the victim (K1v) and the two suspects (K1 and K3). Lanes 3 and 10 are the female fraction DNA from two vaginal swabs (Q1 and Q2) and lanes 4 and 11 are the male fractions from the same vaginal swabs.

RAPE (SUSPECT ELIMINATED)

In an investigation of a series of rapes which occurred on a Virginia military base, two individuals were identified as possible suspects. The suspect was armed with a knife in each case. Military investigators submitted vaginal swabs taken from the victims and known blood samples from suspects to the FBI Laboratory for DNA analysis. The DNA profiles from the semen found on the vaginal swabs did not match the DNA profiles of either of the two suspects (Figure 4).

KIDNAPPING

The unidentified remains of a possible kidnapping victim were discovered several days following the crime. The medical examiner's findings suggested the remains could have come from the kidnapping victim based upon age and sex. DNA analysis was performed on a portion of the remains and compared to the parents of the kidnapping victim. If the remains were the offspring of the parents,

then the DNA profiles from the remains will match a portion of DNA from each parent (Figure 5).

ARMED ROBBERY

An armored car was robbed by two individuals in a northeastern United States city. During the robbery, one of the suspects was wounded by a guard. The stolen getaway vehicle was discovered abandoned a few miles from the crime scene and bloodstains were located in the vehicle. DNA profiles from the bloodstains were compared to the known blood samples from the two suspects who were arrested several months following the robbery. The DNA profiles from the bloodstains matched one of the suspects (Figure 6).

CONCLUSIONS

As has been illustrated by these examples, DNA profiling does provide an excellent technique for potentially individualizing forensic biological stains in a variety of criminal cases. Profiling of

KI K2 K3
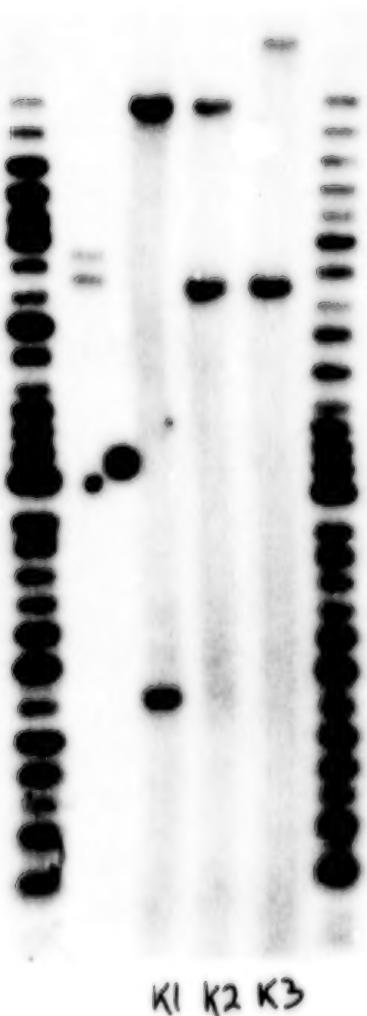
Figure 5. Autoradiogram results from hybridization with single locus probe D1S7. Lanes 1 and 6 are the molecular weight size markers. Lane 2 is the human cell line control. Lanes 3, 4 and 5 are the known blood samples from the suspected mother (K1), the human remains (K2) and the suspected father (K3), respectively.

DNA can provide a clear and definite connection between a suspect and victim in most cases. Further, DNA profiling in some criminal investigations has also been shown to eliminate a falsely associated individual with a particular sample. Lastly, DNA profiling utilized within properly controlled and scientifically sound methods provides a powerful tool for the forensic laboratory and the criminal justice system.
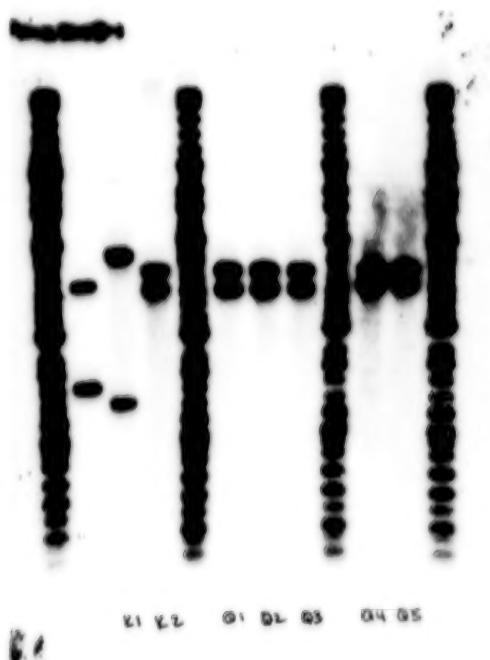
Figure 6. Autoradiogram results from hybridization with single locus probe D2S44. Lanes 1, 5, 9 and 12 contain the molecular weight size markers. Lane 2 is the human cell line control. Lanes 3 and 4 are known blood samples from the two suspects (K1 and K2). Lanes 6, 7, 8, 10 and 11 contain results from the questioned blood stains.

REFERENCES

Adams, D. E. (1988). Validation of the FBI procedure for DNA analysis: a summary, Crime Lab. Digest 15:106–108.

Baechtel, F. S. (1988). Recovery of DNA from human biological specimens, Crime Lab. Digest 15:96–96.

Budowle, B. (1988). The RFLP technique, Crime Lab. Digest 15:97–98.

Budowle, B., Deadman, H. A., Murch, R. S. and Baechtel, F. S. (1988). An introduction to the methods of DNA analysis under investigation in the FBI Laboratory, Crime Lab. Digest 15:8–21.

Monson, K. L. (1988). Semiautomated analysis of DNA autoradiograms, Crime Lab. Digest 15:104–105.

Mudd, J. L. and Presley, L. A. (1988). Quality control in DNA typing: a proposed protocol, Crime Lab. Digest 15:109–113.

BLANK PAGE

DNA ANALYSIS IN THE METRO-DADE POLICE DEPARTMENT CRIME LABORATORY BUREAU

Willard Carl Stuver and Roger Kahn

Metro-Dade Police Department
Miami, Florida

The Metro-Dade Police Department (MDPD) established a DNA analysis program in the summer of 1988 with a start-up budget of $575,000. These funds were allocated to enable the department to hire additional personnel for the Crime Laboratory Bureau, modify existing facilities, develop a training program and put consultants under contract to assist in planning and implementation of the program.

As of June 1989, the Crime Laboratory Bureau has hired a senior level Ph.D. technical coordinator and is in the process of hiring another DNA Specialist and three technicians. Once the staffing is completed, the technical support staff will undergo a training program which will result in an in-house certification. Prior to accepting casework, a study will be undertaken to determine frequencies of the DNA alleles (for 4 genetic systems) that occur within the Metropolitan Dade County area. This sampling is expected to include 300 Blacks, 300 Caucasians and 300 Hispanic individuals. In addition, a study will be initiated to examine our semitropical climate's heat and humidity effect on the integrity and analysis of DNA samples. It is expected that the initial phase of the DNA analysis program will be completed by November 1989, and cases will then be accepted on a limited basis. Before discussing the specifics of the MDPD DNA analysis program, it may be helpful to provide a brief orientation to the location and function of the MDPD Crime Laboratory.

THE MDPD

There are 10 crime laboratories in the State of Florida, 5 local and 5 state. All but one of the local laboratory is located in the Southeast section of the state. The MDPD Crime Laboratory is the largest county laboratory in Florida and provides a full range of forensic services to all law enforcement agencies in Dade County. The laboratory employees 52 full-time staff, 42 of which are technical personnel. The Crime Laboratory serves a population of 1.8 million residents and has a $3.2 million budget. The DNA testing program was made a part of the Biology/Serology Section of the Crime Laboratory Bureau. Not including newly hired DNA personnel, this section employs 5 serologists who process approximately 1,200 cases per year. Sixty percent of these cases are sexual assaults and 30 percent are homicides.

FUNDING

One of the key factors that allowed the Crime Laboratory to initiate a DNA analysis program was the identification of a funding source. The Contraband Forfeiture Statute, which was enacted in Florida in 1985, provides a mechanism for local law enforcement agencies to obtain assets arising from the seizure of contraband materials. Monies from such seizures are maintained in a locally administered Law Enforcement Trust Fund (LETF). In Dade County, LETF money has been used to initiate law enforcement programs which require specialized technical equipment. The DNA testing program met the criteria to be eligible for these funds. The LETF provided $375,000 of the $575,000 DNA budget. These monies were used to purchase equipment and supplies, modify facilities and provide for contractual employees and specialized training. A break down of these expenditures is included in Table 1.

Table 1. FIRST YEAR DNA ANALYSIS PROGRAM COSTS

Expenditure	Allocation
Personnel	$216,000
Equipment	$136,000
Commodities	$98,000
Training	$40,000
Contractual	$85,000
TOTAL	$575,000

PERSONNEL

The bureau recognized that the existing serology staff was heavily burdened with the current caseload and would be unavailable for training in the complex and labor-intensive new methods. Management also realized that some time would have to be devoted to the development and valida-

tion of methods, and that it would be desirable to generate a database of alleles from the local population. These tasks would require substantially more time and effort than existing personnel could provide. Given these considerations, it was decided to hire additional staff. It is, however, the bureau's intention to bring the existing serologists into the program as the opportunity arises.

Special expertise is needed in the forensic application of DNA technology since the analysis methods are still emerging. For this reason, the bureau decided to hire an experienced molecular biologist whose credentials included substantial research and laboratory experience. The bureau also recognized that the introduction of this technology into the courts would require an individual with exceptional credentials and talents. Therefore, the technical coordinator for the program was recruited from a nationally recognized research institution. He was hired in October 1988, as a senior level DNA Specialist. Ultimately, his duties will include the overall responsibility for the technical operations as well as be responsible for developing and implementing new methods. In addition, he will conduct training programs for the laboratory staff, as well as provide informational seminars for the various agencies utilizing the laboratory. He will also be encouraged to establish liaisons with colleagues engaged in DNA testing and research and will be the principal expert presenting testimony on DNA-related issues.

As the DNA testing program has developed so has the need for an additional molecular biologist at the Ph.D. level. This individual will have direct responsibility for supervising the technicians and overseeing the actual testing of the samples. Like the senior DNA Specialist, this individual will interpret results and prepare reports, as well as provide DNA related expert testimony. Additional duties will include implementation of the quality assurance program and the handling and preparation of specialized reagents. In addition to the full-time DNA staff, a one-year consultant position has also been created. This individual will assist in developing new methods or supplementing existing staff when the need arises.

By carefully selecting highly trained DNA personnel to head the project, the bureau has the capability to rapidly evaluate and implement improvements to methods and to provide training resources for the existing serologist as they prepare to work in the DNA Unit. The upper-level staff will provide an in-house training program to the technicians.

To further strengthen the DNA program, a group of molecular biologists from the University of Miami was asked to assist in various aspects of the program's development, from planning to implementation. The bureau envisions its role during implementation to be helping with quality assurance concerns by monitoring proficiency testing. In addition, it will assist in on-going training efforts. Bureau members will serve as independent experts for the courts when necessary. The individuals chosen have backgrounds and expertise ranging from basic molecular biological research to clinical DNA laboratory management.

TRAINING AND CERTIFICATION

All personnel working in the DNA typing program will be required to demonstrate an understanding of the DNA testing, the skill to conduct the testing procedures properly and the ability to understand test results. In order to achieve these goals, a training course is being developed which will provide the equivalent of six credit hours of graduate level course work in molecular biology. In addition to practical laboratory experience, the course includes 20 one-hour lectures covering various aspects of the technology from theory to practice. Trainees will be provided with selected readings, reprints and references. The senior DNA Specialist will give the lectures, with guest lecturers provided by the University of Miami Medical School. Lectures will be available on videotape for future review and study.

Upper level personnel will also be required to demonstrate the ability to interpret test results, to write case reports and to present results to the court. These skills will be sharpened by using mock case materials and courtroom experiences. All facets of the DNA training will be documented by the Section Supervisor. He or she will also have the responsibility for defining the contents of an in-house certification program. The senior DNA Specialist will have the responsibility for conducting the training program and for certifying the trainees. Skills, knowledge and abilities at all levels of performance will be measured through the use of proficiency samples. Both blind and open proficiency testing will be done.

COSTS

A major concern shared by laboratory managers and public administrators is the ability to determine cost of DNA testing and the cost effectiveness of an in-house testing program. To perform

this evaluation, it is important to be able to estimate the numb.. and type of samples that will be processed. We estimate that approximately 30% of sexual assault cases will provide samples for DNA analysis. This number includes only cases with subjects and with sufficient sperm for analysis. Approximately 40% of the homicide cases will likely yield DNA results. This figure represents the percentage of cases with no smoking gun and in which a foreign blood type is present or where blood is found on a subject's clothing. Experience has shown that, on average, there will be 6 samples per case and that a technician can be expected to complete 17–30 samples per week (Personal communications: W. Eubanks, FBI; L. Page, Lifecodes Corp.; M. Stolerow, Cellmark Diagnostics).

At this time, the price of specialized reagents, controls and probes for conducting DNA testing according to the methods developed by the FBI Laboratory at Quantico, VA is approximately $45 per sample. Labor costs add about $55 per sample. The total cost of $100 per sample is substantially less than the $325–$350 charged by commercial laboratories.

An additional advantage of testing in-house is the availability of expert witnesses. Private-sector experts are in great demand, and court schedules must be made well in advance. Since outside experts cost $1,000 per day plus expenses, court proceedings can be very costly. This is particularly evident when the admissibility of DNA evidence is challenged. By having an in-house DNA program the crime laboratory can control costs, evidence, turn-around-time and case prioritization.

SUMMARY

The MDPD has committed substantial resources to the DNA analysis program. At this time, the unit consists of 2 Ph.D. DNA Specialists and 3 technicians and occupies a newly equipped area. The staff is presently engaged in training, methods evaluation and database collection. At the completion of this phase, the laboratory will begin to accept cases. However, the bureau intends to maintain the innovative atmosphere that presently characterizes the unit by encouraging in-house methods development.

BLANK PAGE

DNA ANALYSIS IN THE VIRGINIA BUREAU OF FORENSIC SCIENCE

Paul B. Ferrara

Virginia Bureau of Forensic Sciences
Richmond, Virginia

The Virginia Bureau of Forensic Science began performing DNA analyses on evidential material submitted by the law enforcement agencies of Virginia on May 1, 1989. This marked the culmination of efforts begun over two years ago.

To appreciate the approach to implementation of DNA testing in Virginia calls for a brief explanation of the organization and function of the Virginia Bureau of Forensic Science:

Our single system serves all law enforcement agencies in Virginia—State Police, all local and metropolitan police departments, sheriffs, the Commonwealth's Attorneys (prosecutors) and the State Medical Examiner System.

We serve all these agencies through four laboratories located around the state, employing approximately 160 full-time and part-time scientists and support staff.

Uniquely, the Virginia Bureau of Forensic Science is not organizationally part of any law enforcement agency—rather, it is under the Governor's Secretariat of Administration in the Executive Branch of State Government, specifically an arm of the Department of General Services.

Finally, the Virginia Bureau of Forensic Science system is accredited by the American Society of Crime Laboratory Directors—Laboratory Accreditation board, the only accreditation program for forensic laboratories dealing in forensic evidence.

With that introduction, let me turn to events of two years ago when we developed a more than considerable interest in DNA testing as it applied to forensic materials. So too did the agencies we served. They came to us to seek our advice, information and guidance on the potential application of this new technology to their ongoing investigations. It became very clear that they were going to seek out this available testing from the private sector despite our concerns over the possible prematurity of its use and acceptability in Virginia courts.

So we set out to gain extensive background and in-depth information on the reliability and practical aspects of DNA profiling. The more we investigated, the more our understanding, interest and confidence in the validity of the technology as conducted by the private sector grew. You already have heard others better qualified than I, describe the basis, validity, applicability and reliability of the various DNA testing methods. Suffice it to say, we realized that DNA testing offered revolutionary potential as an adjunct to conventional forensic serological techniques.

Consequently, we felt it then imperative to provide our law enforcement agencies in Virginia with this technology as quickly as practicable, realizing that the sooner we began acquiring this technology, the sooner we could begin to provide proper consultation to our user agencies seeking these services from the private sector. Ultimately this approach would have the further advantage of circumventing the financial and logistical vagaries associated with use of private laboratories for this work and allow the work to be done in-house at our existing state facilities.

Fortuitously, in the spring of 1987, Lifecodes Corp., one of the two private companies providing this type of DNA testing, came to Richmond to discuss and propose to us a one-year technology transfer program. Essentially, Lifecodes would provide four weeks of training for two of our people at their laboratory in Elmsford, NY, to include theory and hands-on instruction. They would provide procedures, quality control of reagents, proficiency samples, supplies and continued technical assistance as necessary. On our part we would commit two fully qualified, experienced forensic serologists with post graduate degrees, full-time for one year to this DNA work. We would pay for the training, purchase requisite equipment, provide dedicated laboratory space and purchase the necessary supplies and reagents from Lifecodes.

So during the period of September–December, 1987 we sought startup funding of approximately $85,000 for equipment/supplies/ training for the period of January–July, 1988 and a budget addendum for another $206,000 for the 1988–1990 biennium. Thus, over 18 months ago, I found myself talking to various boards, legislative committees and commissions in Virginia attempting to impress on them all the advantages, applications and reliability of what was then a technique virtually unheard of among the lay population. Now, many months

later, in Virginia, as a result of highly publicized events during this past year, almost everyone is aware of the value of DNA in crime fighting. This was due, in large part, to the fact that, ironically, at the same time we were soliciting support for a DNA laboratory, a recently released convict by the name of Timothy Spencer, while living in a halfway house, was brutally raping and strangling women in the Northern Virginia city of Arlington, and in the Richmond area. The victims, apparently chosen at random, were of all walks of life and, in the case of the three Richmond area victims, lived on the Southside of the city, giving rise to the designation of the unknown perpetrator as the Southside Strangler. All the investigating agencies had no suspects, no leads and hysteria infected the general populace and the extensive media coverage.

Coincidental with these events, thanks to the support of a progressive Governor Baliles and the foresight of his Secretary of Administration, Carolyn Jefferson-Moss, the Virginia State Crime Commission, our own Department of General Services and the 1988 Virginia General Assembly, the requested resources were provided. On February 1, 1988 we began the first phase of this technology transfer program at Lifecodes.

Meanwhile, an Arlington detective, operating on a long-shot hunch, sent a sample of a suspect's blood along with seminal fluid stains from the Arlington rape/murder to Lifecodes for DNA analysis; on the basis of these results, Timothy Spencer was arrested. Then, in March of 1988, Lifecodes results then linked Spencer to three Richmond area rape/strangulations as well.

On July 8, 1988, a pre-trial motion on admissibility of DNA evidence in the Timothy Spencer case was heard which resulted in a ruling of admissibility. This was the first such ruling in a serial capital murder case in the country. Transcripts of this hearing include the testimony of nationally recognized experts in molecular biology and genetics confirming the validity of this well established technology and its recent application to forensic evidence. In the week following that hearing, Timothy Spencer was convicted by a jury for capital murder and sentenced to die in the electric chair.

Since that time Timothy Spencer has been convicted for the three other slayings and received three additional death sentences. It is important to note that in all these trials, without the DNA evidence, there was insufficient evidence, in the opinion of the prosecutors, to even bring Spencer to trial. In all capital murder convictions, appeals are automatic to the Virginia Supreme Court which

issued a ruling on these cases, accepting DNA evidence.

While all these events were taking place, the establishment of a DNA testing facility in our Tidewater Regional Forensic Laboratory was proceeding on schedule. We converted a 700 ft^2 laboratory area for DNA testing. Our scientists also supplemented their training by taking courses in molecular biology and genetics at Old Dominion University. They completed all phases of the technology transfer program and successfully completed 8 sets of proficiency samples provided by Lifecodes. In addition our scientists conducted two DNA studies on primates and various human tissues, which papers were presented at the American Academy of Forensic Science Meeting in Las Vegas in February, 1989.

All of the work described above was based upon use of the Lifecodes method, using PST-1 as the restriction endonuclease and variable number of tandem repeat probes YNH-24 and V-1.

In January of 1989, we also had the opportunity to participate in the first course offered by the FBI which employed HaeIII as the restriction endonucleases and appropriate probes. Based upon evaluation of this methodology, including factors such as sensitivity, cost, analysis time, availability of supplies and an appropriate population data base, we made the decision to adopt this methodology for use on case materials at the earliest opportunity. The most significant factor in this decision was our desire to foster standardization of DNA methodology in forensic laboratories, especially in view of the fact that the Virginia Bureau of Forensic Science would be the first of many state and local forensic laboratories across the country to implement DNA testing capabilities. Furthermore, participation of our scientists with the FBI's Technical Working Group—DNA Analysis Method (TWGDAM) ensured communication and input to methodologies, quality assurance and data base management.

In March, 1989, Governor Baliles signed into law, effective July 1, 1989, a statute requiring all current and future convicted sex offenders in Virginia to provide a blood sample for DNA profiling. In this st , the Virginia Bureau of Forensic Science is specifically charged with the responsibility of performing these tests and maintaining the resultant DNA Data Bank. As a result of this legislation, we are working closely with the FBI in establishment of a national DNA index. In order to establish a testing facility of sufficient capacity to conduct testing of these samples from convicted

sex offenders, we have submitted a budget request for $1,050,000 over the 90–92 period to provide resources for our central laboratory in Richmond. Until this funding is available, we will store blood samples taken pursuant to this legislation.

Meanwhile, we have promulgated guidelines (Appendix I) to our law enforcement agencies related to acceptance of casework samples for DNA testing. By most standards, our initial operation is small, with a capacity of only about 300 cases per year. Our available manpower consists of three scientists and one technician. Demand for DNA testing will quickly outstrip our capacity. In anticipation of this demand, we have made available space in our new Northern Virginia Regional Forensic Laboratory and in our Western Regional Forensic Laboratory to establish the capacity to at least perform extraction and isolation of DNA from stains at these facilities.

It is important to note that among the first actual case materials submitted for DNA testing on May 1, 1989 was a court-ordered sample for the defense on a case where the prosecution had not opted for DNA analysis and the defense successfully contended that DNA analysis could exonerate his client whereas conventional forensic serology included his client in a population which could have contributed the stains in question.

In conclusion, I hope we have demonstrated in Virginia how a relatively large forensic laboratory system can establish DNA testing in-house and perform such testing in a careful systematic manner using the best available technology and personnel consistent with the guidelines of the Society of Forensic Haemogenetics, the American Association of Blood Banks Parentage Committee and the American Society of Crime Laboratory Directors.

APPENDIX

COMMONWEALTH of VIRGINIA
DEPARTMENT OF GENERAL SERVICES

DIVISION OF
CONSOLIDATED LABORATORY SERVICES

1 NORTH 14TH STREET
RICHMOND, VIRGINIA 23219

BUREAU OF FORENSIC SCIENCE

TO: ALL VIRGINIA LAW ENFORCEMENT AGENCIES

SUBJECT: DNA TESTING GUIDELINES

Effective May 1, 1989, the Bureau of Forensic Science will begin performing DNA analyses on selected suitable evidence which has been submitted to the laboratory for serological examination. DNA testing will, on that date, be available as an additional serological test. Thus, every case submitted to our forensic laboratories for serological examination will be evaluated for applicability of DNA analysis. However, because of our limited capacity to conduct these tests, we must be selective to ensure that our resources are being used most effectively. To that end, the following guidelines have been established for evaluating a case for the potential of DNA analysis:

Guidelines for Performing DNA Analysis on Forensic Samples

1. DNA analysis will generally be limited to cases received after implementation of the technique in the Bureau of Forensic Science.

2. DNA analysis will be limited to suitable probative stains and known blood samples in those cases involving a violent crime, e.g., homicide, sexual assault, aggravated assault. However, due to our limited capacity, priorities will be established by the Bureau of Forensic Science.

3. Samples for DNA analysis must have the necessary conventional serological examinations performed first. The suitability (i.e., size, quality, probative value) of these specimens for DNA analysis can then be determined.

DGS

DIVISION OF ADMINISTRATIVE SERVICES · DIVISION OF ENGINEERING & BUILDINGS
DIVISION OF PURCHASES & SUPPLY · DIVISION OF CONSOLIDATED LABORATORY SERVICES · DIVISION OF RISK MANAGEMENT

4. Known blood samples, collected in a lavender top (EDTA preservative) tube, from the victim and suspect(s) are required for comparison purposes prior to DNA analysis on questioned body fluid stains. Elimination blood samples, e.g., from a victim's husband or boyfriend, may also be required.

5. In certain cases (such as serial homicide/rapes), DNA analysis may be conducted even though a suspect has not been identified.

6. Generally, DNA analysis will be limited to 3 to 4 samples per case, consisting of victim's blood, suspect's blood and 1 or 2 questioned stains.

7. The investigator and the prosecuting attorney may be consulted regarding which specific serological examinations are to be performed on samples too limited for complete evaluations.

8. DNA analysis will only be done for the agencies served by the Virginia Bureau of Forensic Science, unless the Bureau is directed to do otherwise by court order.

9. Sample collection and preservation techniques for DNA analyses are the same as those for any serological evidence.

It must be emphasized that DNA analysis is not a technique that can be performed overnight. Under optimum conditions, the process takes approximately six weeks to complete. Therefore, this should be considered when setting trial dates.

Deanne F. Dabbs, Forensic Serology Section Chief, can be reached at (804) 786-2343 to answer your questions concerning DNA analysis, these guidelines or the status of DNA analysis on your case.

BLANK PAGE

COST, IMPLEMENTATION AND TRAINING FOR DNA ANALYSIS

William G. Eubanks

FBI Laboratory
Washington, D.C.

The implementation of Deoxyribonucleic Acid (DNA) analysis in the FBI Laboratory's DNA Analysis Unit (DNAU) in October 1988, was the culmination of over two years of research and development by scientists at the Forensic Science Research and Training Center (FSRTC), Quantico, Virginia. The research was directed toward the development and validation of a restriction fragment length polymorphisms (RFLP) protocol as well as studies to determine the effects of age, environmental conditions, contaminants, substrates, mixing of body fluids and other factors on the ability to reliably apply the RFLP method to forensic evidence. Following a period of initial testing to insure that all phases of the method were operating properly, the DNAU began applying this technology to casework in December 1988. The commitment of resources and funds for this ambitious project were unprecedented in FBI Laboratory history and were the result of the support of FBI Director William S. Sessions and FBI Laboratory management.

Although limited research of DNA technology was already underway at the FSRTC, recognition of the unparalleled potential of this technology in the investigation of violent crimes led to the formation of a DNA research team in July 1987. The FSRTC scientists were augmented by additional special agent and technician personnel from the Scientific Analysis Section who were detailed to the research team. As a result of the rapid progress of the research, a DNA Implementation committee was formed in early 1988 which consisted of management and technical personnel from the Laboratory Division as well as the Legal Counsel, Technical Services, Training, Records Management and other FBI Headquarters (FBIHQ) Divisions. This committee was tasked with all aspects of planning for the implementation of DNA analysis at FBIHQ and the rapid transfer of the technology to FBI as well as State and local crime laboratory personnel. Major considerations in implementation at FBIHQ included location and renovation of suitable laboratory space, purchase of equipment and supplies, personnel enhancement, development of examination protocols with the Serology Unit, training, funding and formulation of a DNA case acceptance policy.

This paper will discuss the experience of the FBI Laboratory in implementing DNA analysis at FBIHQ as well as training requirements and DNA cost estimates. Hopefully this information will be of assistance to other crime laboratories planning to implement DNA technology.

LABORATORY SPACE

With the relocation of the Metallurgy Unit to other space, approximately 2,200 ft^2 of space was designated for the DNAU (Figure 1). Fortunately this area already had separate rooms for the specialized aspects of DNA analysis such as autoradiography and probe labeling. Therefore, any major structural changes were precluded. Specialized laboratory furniture was installed in order to efficiently perform the technique. The two largest work areas are used for sample preparation and extraction, electrophoresis, Southern blot, hybridization, various washes, membrane stripping and other processes. This area includes two large 4 x 11 ft benches suitable for eight technicians, a double fume hood used in the extraction process, a biological contaminant hood and space for double door freezers and refrigerators. There are additional benches along the walls for other equipment.

A small dark room for autoradiography contains a red safe light for handling of films, an automatic x-ray processor and an x-ray copying machine. It was necessary to install a revolving dark room door at the entrance.

The reagent preparation room contains bench space along the walls for instruments used in reagent preparation, cabinets for glassware, a fume hood, regular and distilled water service and a wall-mounted multi-staged deionized water system fed by distilled water. A large autoclave, essential in DNA analysis, is located in another area of the FBI Laboratory.

A separate room contains two computer image analysis work stations used in the semiautomated analysis of DNA autorads. Viewing and photographing of yield, post restriction and analytical gels under ultraviolet light is performed in another small room adjacent to the main work area.
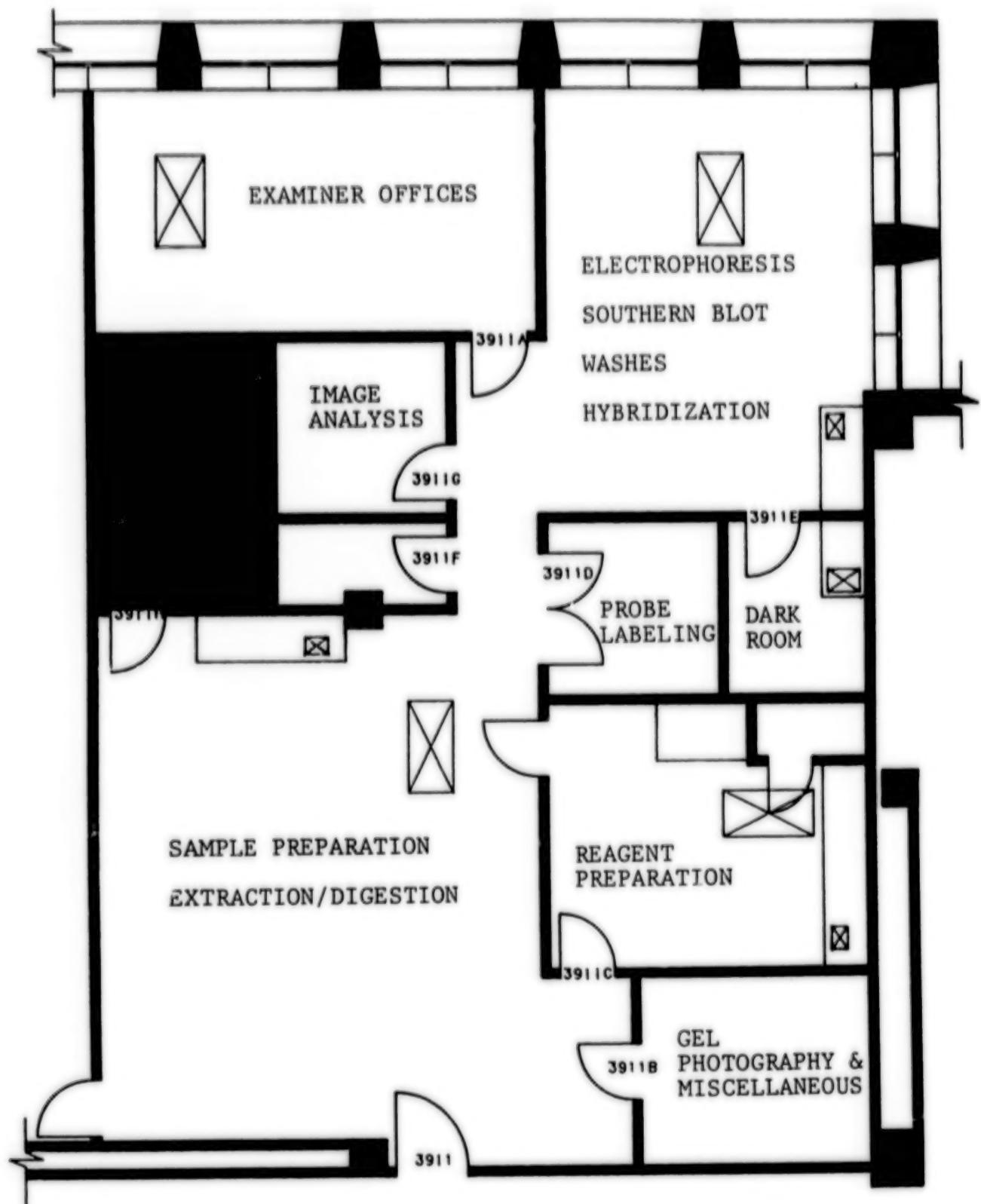
Figure 1. Space configuration for the DNA Analysis Unit. The rectangles with an X indicate access beneath the floor.

The probe labeling room is a small isolated space adjacent to the main work area. It contains a hood dedicated solely to the preparation of radio-actively-labelled DNA probes. This room also contains a sink with tap water, a small freezer for storage of radioisotope and sufficient bench top space.

190

The remaining DNAU space is for examiner offices and computer support equipment.

EQUIPMENT AND SUPPLIES

The conduct of RFLP DNA analysis does not require any sophisticated equipment items. Laboratories that currently perform electrophoretic genetic marker analysis on body fluid stains will likely have some of the items listed in Table 1. It should be noted that most of the electrophoretic equipment needed is designed specifically for DNA analysis. Other necessary equipment items commonly found in laboratories include a pH meter, stirring hotplate(s), balances, glassware, aliquot mixer(s), refrigerators, freezers and an ice machine.

Table 1. DNA EQUIPMENT

Equipment Item	Approximate Cost/Item
Autoclave (bench top)	$4,600
*UV Spectrophotometer	12,000
UV Transilluminator	2,300
Electrophoresis power supply (2)	2,500
Electrophoresis tanks (4)	400
Microfuge (general use)	1,200
Microfuge (dedicated to isotopes)	1,200
Microliter pipettor (variable range) (3)	250
Water bath (3)	2,000
*Vacuum centrifuge	4,300
Ultra-low freezer	6,300
*Environmental rotary shaker	5,100
*Vacuum oven	1,000
Gel photography equipment	1,000
*X-ray film autoprocessor	7,100
Platform shaker	1,000
Benchtop radioisotope counter	2,400
Radioactivity survey meter	250
Intensifying screens (pair) (5 pair)	300

The items with an asterisk (*) are considered desirable but not essential for performing DNA typing.

As previously discussed, several types of laboratory hoods are necessary to safely conduct DNA typing. A fume hood for the handling of organic solvents such as phenol and chloroform is essential. Also, a specialized hood located in an isolated area for the handling of radioactively labeled probes is necessary. Additionally, a laminar flow hood is recommended to protect personnel from biohazards when drying out a liquid blood sample.

The DNA laboratory will also need an autoclave for sterilizing glassware. Depending on the size of the laboratory, it may be necessary to purchase an often expensive, large autoclave. Another essential item is a multi-stage deionized water system connected to the laboratory's distilled water.

Analysis of RFLP requires a considerable inventory of supplies. Some of the supply items, such as DNA probes, restriction enzymes and radioisotopes, are quite expensive. However, other items, such as microcentrifuge tubes and micropipette tips, are used in large quantities, though they are not expensive individually. There are also a significant number of chemical reagents used. The DNAU also utilizes in its RFLP procedure a molecular weight size marker, human cell line control, combination marker, visual marker and yield gel calibration standards. These controls are essential for quality control purposes and are expensive. A more detailed explanation of DNA costs is presented later in this paper.

CASE ACCEPTANCE POLICY

A DNA case acceptance policy was implemented to insure that this new technology was applied only to probative evidence in violent crime cases. This policy was necessary because of the amount of time required to conduct the RFLP procedure, expected increase in case submissions and the limited resources of the FBI Laboratory. Probative evidence is defined as body fluid stains on items in which DNA analysis could potentially associate the suspect with the victim or the suspect with the crime scene.

In general, this policy states that the FBI Laboratory will accept evidence for DNA analysis from current, violent personal crimes where appropriate standards for comparison are available. DNA analysis on State and local cases is limited to homicide, rape and serious aggravated assault cases in which a suspect has been identified. In certain cases such as serial homicide/rape and sexual assaults of young children, evidence is accepted even though a suspect has not been identified. A known blood sample from the suspect and victim is required for comparison purposes. In general, cases with crime dates prior to December 1988, the month the FBI Laboratory began using DNA analysis in casework, are referred to one of the private DNA testing laboratories.

State and local law enforcement agencies are encouraged to submit body fluid stained evidence from violent crimes to their local crime laboratory for preliminary serological testing. If stains of potential probative value are identified which are suitable in quality and quantity for DNA testing, and if the laboratory does not have DNA capability, the FBI Laboratory will accept the evidence samples contingent on the case which meets acceptance guidelines. Thus far more than half the DNA cases

received have been of this type. The law enforcement agency should discuss the case with the DNAU prior to submitting the evidence.

COORDINATION WITH THE SEROLOGY UNIT EXAMINATION PROTOCOLS

The implementation of DNA analysis in casework requires close coordination and communication between examiners in the DNAU and Serology Unit (SU) in order to facilitate the logical and smooth flow of evidence. Examiners in both units are involved in the assessment of evidence as probative or non-probative based on information supplied by the contributor, quality and quantity of stain material available for DNA analysis, and preliminary serological and/or genetic marker results. All violent crime cases in which a suspect has been identified and known standards are available are considered potential DNA cases.

Probative human bloodstains of sufficient quality and quantity are screened electrophoretically prior to submission to the DNAU. Genetic marker screening normally includes subtyping of PGM and possibly Gc depending on the amount of available stain. Probative bloodstains of insufficient quantity for preliminary genetic marker screening are normally submitted directly for DNA analysis. Examples of non-probative bloodstains are the victim's blood on the victim's clothes or at the scene, or the suspect's blood on the suspect's clothes. Only cuttings or scrapings from body fluid stains as well as swatches from the known blood samples dried out on cotton sheeting are submitted for DNA analysis. All probative semen stains are forwarded to the DNAU for analysis. The DNAU examiner further evaluates the stain(s) for DNA analysis based on its quality, quantity and potential probative value.

Body cavity swabs (1 to 3) from a single orifice are combined as practical with one-half of the swab(s) being analyzed for prostatic antigen (P-30) and ABH antigens and the remaining halves (dry, unextracted) submitted for DNA analysis. Vaginal smear slides are examined microscopically for spermatozoa and if found, the dry unstained slide is submitted for DNA analysis. However, experience has shown that the swab is the best evidence for DNA typing.

In general, a bloodstain at least the size of a quarter and a semen stain the size of a dime are considered adequate for DNA analysis. However, DNA profiles are frequently obtained from smaller quantity samples. Obviously there are a number of other factors such as stain substrate, concentration, age and environmental conditions which will affect a stain's suitability for DNA profiling.

OPERATING EXPENSES

As of this writing, the DNAU had been operational for 25 weeks. Approximately 300 cases have been received consisting of over 1,400 samples. The average case consists of about 5 samples. Based on monthly operating costs to perform DNA analysis in the FBI Laboratory, the estimated annual cost for supplies is $271,500 (Table 2). This figure includes the cost of probes, restriction enzyme, isotope, controls and reagents/other supplies. It is projected that a fully staffed unit can process about 10,000 samples in a year. Thus the cost per sample, not including labor, is approximately $27.15.

Table 2. ESTIMATED OPERATING EXPENSES

PRODUCTIVITY:
- One technician can extract 30 samples/week; seven technicians can extract 210/week.
- 7 samples/gel.
- 4 membranes (28 samples)/hybridization.
- 48 week work year.
- Therefore, ~10,000 samples/year can be processed.

OPERATING COST/MONTH:

3 Lifecodes probes labeled $30 \times 10 \wedge 6$cpm (D2S44, D168S85 and D17S79)	$3,100.00
Ladder probe	10,500.00
Isotope	600.00
Cellmark D1S7	900.00
Other supplies/reagents	~3,000.00
Total	18,100.00

$18,100/month × 12 months = $217,200
Since the unit is not at full production, an increase of 25% would bring the total estimated operating cost per year to $271,500.

COST PER SAMPLE:
$271,500/10,000 samples per year = $27.15 per sample. This figure does not include the cost of labor.

As previously discussed, the above costs are operational only and do not include the initial cost of establishing a forensic DNA testing laboratory. It cost approximately $150,000 to completely equip the DNAU. However, because of the FBI's caseload, it is necessary in some instances to purchase 2 or more of the same equipment item which may not be necessary for a smaller laboratory. An additional $50,000 was spent for supplies to get started. These figures do not include any costs for space renovation or laboratory furniture.

TRAINING

Selection of personnel for DNA analysis is an important aspect to be considered when planning for implementation. The scientists who will be interpreting the test data and testifying should have an advanced degree (Masters or higher) in a bio-

logical science. The fields which are most relevant to this technology are molecular biology and genetics. Prior experience in forensic serology is highly desirable.

The FBI's training program for special agent examiners requires the trainee to complete the one year of training for qualification in forensic serology and is about 4 months of training on DNA analysis. This training encompasses an intensive 4 week DNA course at the Forensic Science Research and Training Center, FBI Academy which stresses the theoretical understanding of the basic concepts underlying this technology. Also included in the training is practical application of the techniques at the laboratory bench. The examiner is also required to thoroughly review the relevant scientific literature which includes various scientific journal articles and textbook materials. The examiner supplements his study with approximately 8 weeks of training at the laboratory bench in which he becomes thoroughly competent at performing all steps of the procedure. Examiners are encouraged to take courses in molecular biology to further enhance their knowledge and credibility as experts in the field. Finally, the examiner is required to successfully complete proficiency tests with no errors and to demonstrate his knowledge and expertise in a series of oral boards and moot courts.

SUMMARY

The implementation of forensic DNA testing in a crime laboratory requires careful planning on the part of laboratory managers, a commitment to the necessary resources and funding for the equipment and supplies. The laboratory may also incur additional significant costs for space renovation or acquisition of new space. Also important is coordination of examination protocols between DNA testing and serological analysis. The selection of qualified personnel and a formal training program are mandatory. Despite the problems which may be encountered in the implementation of this new technology, the potential of forensic DNA testing in the solution of violent crimes will result in a more efficient and effective law enforcement system.

BLANK PAGE

LEGAL ISSUES PERTINENT TO TYPING OF DNA

George W. Clarke

Office of the District Attorney
San Diego, California

The results of scientific testing of physical evidence in criminal cases have been presented in courts of the United States for over one hundred years. Scientific testimony has frequently played a significant role in the conviction of individuals responsible for the commission of both violent and nonviolent offenses. Equally significant, scientific evidence has repeatedly resulted in the exoneration of potential suspects in criminal offenses. DNA identification testing methods present the forensic examiner with a modern and powerful tool for the resolution of such crime.

THE FRYE RULE

The standard for admission of results of testing of physical evidence utilizing new or novel scientific techniques was set forth in 1923 by the District of Columbia Circuit of the United States Court of Appeals. In that decision the federal court concluded:

> "Just when a scientific principle or discovery crosses the line between the experimental and demonstrable stages is difficult to define. Somewhere in this twilight zone the evidential force of the principle must be recognized, and while courts will go a long way in admitting expert testimony deduced from a well-recognized scientific principle or discovery, the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs." [1]

This rule requiring general acceptance of a new or novel scientific technique—commonly referred to as the Frye rule—was utilized in 1923 by the above federal appeals court to uphold a ruling by a trial court disallowing a defendant's offer of results of a polygraph examination. While extensive criticism of the Frye rule has been leveled by courts and commentators for the past 65 years, it has remained the governing standard for both the federal courts and approximately 45 states of this country.

The California Supreme Court has noted what it believes to be compelling reasons for the Frye rule and its application to scientific testing methods:

"Lay jurors tend to give considerable weight to scientific evidence when presented by 'experts' with impressive credentials. We have acknowledged the existence of a '. . . misleading aura of certainty which often envelops a new scientific process, obscuring its currently experimental nature.' [Citation omitted.] '. . . scientific proof may in some instances assume a posture of mystic infallibility in the eyes of a jury . . .' [Citation omitted.]" [2]

The Relevant Community

The establishment of general acceptance in the particular field in which the scientific technique belongs depends in large part on the manner in which the relevant community is defined. Experience in the area of serological protein examination through the use of electrophoresis is particularly instructive. The extent of the relevant community from which the court will entertain evidence in the course of establishment of general acceptance has varied from the restrictive view undertaken by the Supreme Court of the State of Michigan to more expansive views set forth by other courts, including more recently the state of California.

The Michigan Supreme Court concluded in 1986 that the multi-system technique of electrophoretic analysis of evidentiary stain material was not shown to have gained general acceptance in the relevant community. [3] The Michigan court, upon review of the trial court record of testimony of numerous expert witnesses, found that only two individuals presented testimony from which the issue of acceptance could be determined.

Those two witnesses—Dr. George Sensabaugh of the University of California at Berkeley and a private consultant, Dr. Benjamin Grunbaum—were concluded by the Michigan court to be the only expert witnesses free of interest and partiality whose testimony was presented the trial court. Rejected by the Michigan court was the testimony of numerous other experts, including highly trained and experienced serologists.

A contrary view—and that subscribed to by the majority of jurisdictions in the United States—allows and considers the testimony of witnesses

[1] Frye v. United States (D.C. Cir. 1923) 293 Fed. 1013, 1014.

[2] People v. Kelly (Calif. 1976) 17 Cal.3d 24, 31–32.
[3] People v. Young (Mich. 1986) 391 N.W.2d 270.

with sufficient qualification and foundation for the rendition of expert opinions about a scientific methodology or technique. While acknowledging that a witness whose opinions are to be considered on the issue of general acceptance must not be so personally invested in establishing the technique's acceptance that he might not be objective about disagreements in the community,[4] California has concluded:

> "It would be a strange perversion of Kelly/Frye to exclude the opinions of analysts, at least well-credentialed ones. Those who work closest to a technique may be uniquely aware of inherent reliability problems." [5]

The opinions of casework forensic analysts obviously provide the court with testimony and evidence from which the issue of general acceptance can be more considerately resolved.

General Acceptance

The establishment of general acceptance of a new or novel scientific technique, as noted previously, was found by the Frye court and subsequent court opinions to ensure that a sufficient foundation exists prior to the introduction of results of such testing methods. However,

> "The Frye test does not demand the impossible—proof of an absolute unanimity of views in the scientific community before a new technique will be deemed reliable; any such unanimity would be highly unusual; . . . Rather, the test is met if use of that technique is supported by a clear majority of the members of that community." [6]

The requirements of the Frye rule do not include a determination by the trial court of reliability of a scientific technique.[7] Rather, the court is explicitly limited to a conclusion from the presentation of evidence whether, following a fair overview of the matter, scientists significant either in number or expertise publicly oppose a technique as unreliable.[8]

The nature of evidence presented a trial court for purposes of demonstrating general acceptance can obviously vary. That evidence usually includes the testimony of expert witnesses with theoretical understanding or knowledge and actual experience with the technique, as well as forensic scientists who have experience using the technique in the examination of physical evidence. Further evidence presented may include scientific publications,

whose use is limited to assisting the court in the determination whether the technique is accepted in the community.[9]

THE KELLY/FRYE RULE

California and several other jurisdictions impose a secondary requirement for admission of results of testing utilizing new or novel scientific techniques. That additional prerequisite requires the proponent of the evidence to demonstrate that correct scientific procedures were utilized in the testing at issue.[10]

The question whether correct scientific procedures requires establishment that the testing was performed properly at each stage—or that simply the technique actually used in the case is that shown to have been generally accepted—now appears settled, at least in California. The California Supreme Court has recently concluded the latter view applies:

> ". . . the Kelly-Frye rule tests the fundamental validity of a new scientific methodology, not the degree of professionalism with which it is applied. [Citation omitted.] Careless testing affects the weight of the evidence and not its admissibility, and must be attacked on cross-examination or by other expert testimony." [11]

Application of the Frye Rule to DNA Identification Testing

DNA identification testing has been and will continue to be contested in courts of the United States at both the Frye hearing and trial levels. Until such time as published appellate court opinions or other precedent accepted by trial courts eliminates the need for a Frye hearing in jurisdictions of this country, fulfillment of the requirements of Frye will be necessary upon entry of a timely objection by an opponent party.

General acceptance of the scientific theory underlying DNA identification testing appears to be without controversy among the scientific community. Scientific research dating to the work of Watson and Crick has amply demonstrated the structure of the DNA molecule and individualization of the human genome—with the lone exception of identical twins.

Similarly, individual techniques have been demonstrated to be reproducible and reliable through the application of molecular biological theory to research, clinical and forensic settings.

[4] People v. Reilly (Calif. 1987) 196 Cal.App.3d 1127, 1138.
[5] Id., at p. 1140.
[6] People v. Guerra (Calif. 1984) 37 Cal.3d 385, 418.
[7] People v. Shirley (Calif. 1982) 31 Cal.3d 18, 55.
[8] People v. Brown (Calif. 1985) 40 Cal.3d 512, 533; People v. Shirley, supra, at p. 55.

[9] People v. Kelly, supra at p. 35; People v. Shirley, supra, at p. 56.
[10] People v. Kelly, supra, at p. 30.
[11] People v. Farmer (Calif. 1989) 47 Cal.3d 888, 913.

While the emphasis of forensic use of DNA testing has centered around the restriction fragment length polymorphism (RFLP) technique, other methodologies have been or will be the subject of Frye inquiry. These approaches include the polymerase chain reaction and dot-blot analysis method, direct and automated DNA sequencing, and the examination of mitochondrial DNA.

General Acceptance and Validation

The relationship between general acceptance and validation of a scientific testing method has been subject to differing interpretations in courts of this country. Little controversy appears to surround the significance of validation in demonstrating the ability of a method to produce valid and reliable results. Rather, disagreement surrounds the questions of who must conduct any such validation testing and whether the sufficiency of that validation is to be determined by the court in the course of a Frye hearing or by the relevant community of scientists.

The Michigan Supreme Court concluded that independent verification was required by the Frye rule and that developers of a particular scientific technique—in that instance, the multi-system electrophoretic method—could not provide the validation necessary to establish general acceptance.[12] Specifically, the Michigan court noted:

> "The scientific tradition expects independent verification of new procedures. When other scientists analyze and repeat the tests, they counteract the dangers of biased reporting. It is scientists not responsible for the original research that confirm its validity."[13]

Amazingly, the Michigan court disregarded the testimony of Special Agent James Kearney regarding extensive validation and casework use by the Federal Bureau of Investigation of the multi-system technique.[14]

This restrictive view of validation adopted by the Michigan court can be contrasted with the experience of other states. California has recently noted that Michigan stands alone in its stance with regard to independent verification.[15]

Furthermore. guidelines of the California Association of Criminalists with regard to forensic serology are illuminating. Those guidelines conclude that validation of serological methods can occur through a history of use and acceptance in the particular field, description and appearance in the appropriate scientific literature, or developer validation through blind trial testing, collaborative proficiency testing, or split sample testing.[16] Guidelines for validation of DNA testing methods in the forensic laboratory have already been described (see, for example, Budowle et al. 1988).

Degradation and Contamination

Attacks upon forensic serology in the past ten years have regularly included allegations that despite the reliability and acceptance of various techniques for the analysis of whole blood and other bodily fluids, usage of these methodologies for the analysis of evidentiary stain material is unreliable. The Michigan Supreme Court concluded the allegations were founded and included that ground as a basis for its rejection of the multi-system electrophoretic technique.[17]

Again, rejection of the Michigan court's conclusions has occurred in other jurisdictions. Citing studies undertaken by the scientific community, the existence of limitations on the very ability of proteins to undergo apparent transformations of phenotypes—described as protein chemistry at times—and repeated concessions even by the primary opponent of admissibility of evidentiary stain electrophoretic results, the contention was dismissed.[18]

The allegation that degradation and contamination are applicable to evidentiary materials tested by DNA identification methods has already been raised in the United States in the course of Frye admissibility hearings. The existence of a DNA chemistry—akin to that previously discussed with regard to proteins—preventing the appearance of false positive results has been shown. Avoidance of erroneous results is thus possible (Kanter et al. 1986).

Population Frequency Distributions

The individualization of unknown evidentiary material through DNA identification testing techniques is possible through the existence of population frequency distribution statistics. Those statistics provide the foundation for conclusions with regard to relative frequency of base pair sequences

[12] People v. Young, supra, at p. 283.
[13] Id.
[14] People v. Young, supra, Reporter's Transcript of Proceedings, May 24, 1984, pp. 314-323.
[15] People v. Morris (Calif. 1988) 199 Cal.App.3d 377, 389-390.

[16] California Association of Criminalists & California Department of Justice, "Report of a Symposium on the Practice of Forensic Serology" (1987) p. 51
[17] People v. Young, supra, at pp. 281-283.
[18] See, for example, People v. Partee (Ill. 1987) 511 N.E.2d 1165, 1185; People v. Reilly, supra, at pp. 1149-1151; People v. Morris, supra, at pp. 388-390; State v. Washington (Kan. 1981) 622 P.2d 986, 990; see also Budowle, B. and Allen, R.C. (1987) Electrophoresis reliability: I. The contaminant issue J. Foren. Sci. 32:1537-1550.

common to evidentiary material and known individuals (see, for example, Jeffreys *et al.* 1985).

Initially, the question must be raised whether a foundation for the validity of population frequency distribution data is even necessary in the context of a Frye hearing. Frye, as previously discussed, addresses itself to general acceptance of a new or novel scientific method. No reference is made to data utilized as a result of population genetic surveys for purposes of rendering conclusions about relative frequency of the appearance of characteristics revealed by testing. In fact, courts have normally separated the issues of technique acceptance and frequency statements, even in the context of forensic serology.[19]

One exception to this concept of separation of general acceptance and frequency distribution conclusions exists in the only currently published appellate court review of the admission of DNA testing evidence. A Florida court of appeal, while finding general acceptance of DNA testing of evidence material, also concluded that the use of population frequency distribution statistics was generally accepted.[20] However, note should be made that Florida does not subscribe to the Frye rule, but rather determines admissibility of results of the use of new scientific techniques on the basis of a relevancy standard.[21]

Thus, admissibility of population frequency statistics with regard to DNA testing results should not require a foundation within the Frye rule. As was noted by the California Supreme Court in 1985 in a case in which it found a sufficient showing of general acceptance of electrophoretic analysis of evidence stain material had not been presented,

> ". . . both California and the majority of other jurisdictions have traditionally admitted statistical blood group evidence of this kind in criminal cases, even where it simply includes the accused within the class of possible donors." [22]

DNA Profile Information Collection

The potential for the examination, collection and creation of databases for DNA profiles of known individuals has already been realized. Statutes in numerous states already provide for the testing of blood or other bodily fluids of convicted felons and maintenance of such information in state information files. In California, for example, a stat-

ute requires all persons convicted of any of numerous sex offenses to provide samples of blood and saliva for purposes of genetic marker testing and data maintenance.[23]

The potential for resolution of violent offenses is obvious. Through examination of evidentiary material of unknown origin and comparison to existing DNA profiles of known individuals, identification of potential suspects is possible. The benefits to the community and law enforcement have been described (Hicks 1988).

Concerns about the impact of such data collection upon rights of privacy have already been voiced. However, with respect to DNA profiles of those persons who have already been convicted of crime, any expectation of privacy as to that individual's genetic marker profile is clearly reduced and not subject to traditional Fourth Amendment protection.[24] The maintenance of DNA profiles of even persons simply arrested can be likened to long-standing policies regarding booking photographs and traditional fingerprints.

CONCLUSION

The ability of DNA identification testing techniques to provide individualizing information about evidentiary stain material of otherwise unknown origin is already known. The importance of proper presentation of such evidence in court in the course of criminal prosecutions is obvious. Only through clear and concise presentation by both attorney and scientific expert may the results of any scientific test be conveyed to court and jury. Establishment of the foundational requirements of the Frye rule will clearly serve to enhance the presentation of relevant evidence by the proponent of DNA testing results.

REFERENCES

Budowle, B., Deadman, H., Murch, R. and Baechtel, F. (1988). An introduction to the methods of DNA analysis under investigation in the FBI Laboratory, Crime Lab. Dig. 15:18–20.

Hicks, J. (1988). DNA profiling—a tool for law enforcement, FBI Law Enf. Bull. Aug. 1988.

Jeffreys, A., Wilson, V. and Thein, S. (1985). Individual specific "fingerprints" of human DNA, Nature 316:76–78.

---

[19] See, for example, People v. Brown, supra, at p. 536; People v. Morris, supra, at p. 391.
[20] Andrews v. State (Fla. 1988) 533 So.2d 841, 850.
[21] Id.
[22] People v. Brown, supra, at p. 537, fn. 6.

[23] California Penal Code, section 290.2.
[24] People v. Thomas (Calif. 1975) 45 Cal.App.3d 749, 756; People v. Ott (Calif. 1978) 84 Cal.App.3d 118, 124–125.

Kanter, E., Baird, M., Shaler, R. and Balazs, I. (1986). Analysis of restriction fragment length polymorphisms in deoxyribonucleic acid (DNA) recovered from dried bloodstains, J. Foren. Sci. 31:407.

BLANK PAGE

SUMMARY OF THE INTERNATIONAL SYMPOSIUM ON THE FORENSIC ASPECTS OF DNA ANALYSIS

John W. Hicks

FBI Laboratory
Washington, D.C.

This paper is a summary of what was a highly successful symposium. To do that, I will review the FBI program for DNA implementation and discuss, in particular, the future development of DNA indexes.

DR. ALEC JEFFREYS AND THE NARBOROUGH CASE

First let us review the case mentioned by Dr. Alec Jeffreys (Jeffreys *et al.* 1990), the case which found Colin Pitchfork guilty of the 1983 rape-murder of 15 year-old Linda Mann and the 1986 rape-murder of another young woman, Dawn Ashworth. The application of DNA technology in this case captured the interest and imagination of the media, the public and law enforcement worldwide. Investigators quickly recognized the potential impact of this powerful new investigative tool in helping them to solve some of the most difficult kinds of cases they handle. The case is instructive for us in that it illustrates how this tool might best be applied and provides a blueprint for what we in law enforcement must do to realize its full potential.

The Pitchfork case demonstrated the utility of DNA technology in the important four following applications:

1. The technology was found essential in linking the individual crimes thus enabling the police to bring focus to their investigative activities.

2. DNA testing established the innocence of the prime suspect, at least as far as his participation in the sexual assault.

3. The results of the DNA test on the crime scene evidence was used to establish investigative direction which in this case involved the initiation of a voluntary mass-screening program.

4. The technology was used to conclusively identify the perpetrator of the crimes.

CRIME IN THE UNITED STATES

Keeping these applications in mind, let us now examine the crime picture in the United States—particularly violent crimes. According to statistics found in the FBI's Uniform Crime Reports, there are in excess of 90,000 reported rape cases in the United States each year. Only about 50% of these cases are closed with the arrest of a subject. Other studies have estimated that as many as half of the rapes which occur in this country are not reported by the victims. There are 20,000 homicides per year in the United States, 30% of which are unsolved. This is a significant crime problem on which DNA technology can be brought to bear.

According to a study released by the Bureau of Justice Statistics earlier this year, there is a high recidivism rate among convicted criminals. Their survey showed that of over 100,000 convicts released from prisons in eleven states in 1983, more than 62% were rearrested within three years. Violent offenders were more likely to be rearrested than those which had been convicted of non-violent crimes. These individuals were mobile—1 of 8 of those rearrested were charged in a state other than the state in which they were originally convicted.

While this represents a picture of the violent crime problem in the United States, there may be some similarities to such crimes in Europe. As Brian Parkin (Parkin 1990) of the Metropolitan Police Laboratory in London mentioned, initiatives to unite the European Economic Community may also permit greater mobility of certain types of criminal offenders between countries. This will mean a greater need for cooperation, coordination and exchange among police agencies in those countries.

The work done by the psychological profilers at the FBI's National Center for the Analysis of Violent Crime has shown that the individuals who commit these types of crimes are predictable. Once their criminal behavior is established, they continue to commit their offenses. They move from jurisdiction to jurisdiction committing their crimes on randomly selected victims, and thus are difficult for police to identify. Intuitively, it is clear that many of the unsolved cases I mentioned earlier could be resolved if we were able to definitely link these crimes and tie them to a particular offender through the application of DNA testing.

THE FBI PROGRAM

There have been numerous presentations with a view of FBI activities which have been underway for several years now (Adams *et al.* 1990; Baechtel 1990; Budowle 1990; Budowle and Monson 1990; Eubanks 1990; Monson and Budowle 1990). These activities have been designed to bring DNA technology into full use in our criminal justice system in a deliberate and responsible way. Let me now summarize the four basic phases of the FBI's program.

The first phase, research and development, was carried out by our staff of research scientists at the FBI Laboratory's Forensic Science Research and Training Center (FSRTC). They went through a process of learning and assessing DNA test methods, refining the methods for application in forensic cases and insuring that our methods were developed and thoroughly evaluated so that they could withstand the scrutiny of the courts. We wanted to satisfy ourselves that our DNA test would provide consistent and reliable results, could be easily interpreted by the forensic analyst and would provide information which could be understood in the courtroom. Various controls were incorporated in the FBI's DNA test protocol in consideration of these needs and requirements. Undoubtedly, as we move forward with this technology we will identify additional controls to be used on the process. We are working now on an internal lane standard to control for possible band shifting in the electrophoresis of DNA. The need for this control was stimulated by technical discussions held at the Banbury Center (Cold Spring Harbor Laboratory, NY) last November and reinforced by recent Frye proceedings in the Castro case in New York City. I am confident that our DNA test protocol is ready for the courts but we must continually reexamine our procedures and be alert to refinements as we apply this very dynamic technology.

The implementation phase, the second phase, involved moving the technology from the research laboratory to the operational setting at our main laboratory in Washington. We quickly overcame minor problems with equipment and underwent a proficiency program to assure the tests could be readily transferred. We have been pleasantly surprised with the efficiency of the DNA test. Unlike the high rate of inconclusive calls we have experienced with classical protein typing systems, we have successfully typed DNA in well over 80% of the samples tested. We have been able to resolve numerous evidence situations which, heretofore, could not be resolved. With each new case our examiner personnel are gaining increased appreciation for the tremendous power of DNA technology to provide important investigative information.

We are in the midst of the technology transfer phase of our program, the third phase. The primary mission for our FSRTC is to reach out to the forensic testing community and to facilitate their implementation of DNA testing through training programs and other support. I am especially proud of the Visiting Scientist Program. It has proven mutually beneficial to the FBI DNA research effort and the state and local laboratories and is an excellent example of cooperation at its best. We have demonstrated through this program and the 4-week DNA training course that the test methods can be readily transferred to crime laboratories throughout the United States.

The fourth phase of the FBI program involves the establishment of specific types of DNA indexes. It is important that these indexes be narrowly defined to meet specific requirements in particular applications.

There are two types of DNA indexes under development: statistical and investigative.

The statistical index is necessary for laboratory personnel to interpret the significance of a DNA test result. It contains no names or other individual identifiers. The file contains the DNA types of several hundred individuals randomly collected from a population. The DNA types are grouped according to racial or ethnic origin. Using this information the forensic analyst is able to determine the estimated frequency of occurrence of a particular DNA type. Although not yet computerized, a statistical DNA index is now in use in the FBI's DNA Unit and the index will be made available to any crime laboratory which uses the FBI's DNA test procedure. As more laboratories come on line with DNA typing and do local population studies, we expect this index to grow rapidly.

Investigative indexes will contain individual and case-related DNA profiles and will be of several types to support specific investigative needs:

Known offender files will contain DNA profiles of individuals convicted of specific crimes, typically rape, homicide or aggravated assault. These files will be collected at the state level under the authority of individual state legislation. We expect that the first states to establish such DNA indexes will be Virginia, Colorado, Washington and California. Based on our discussions with officials in these states, the known offender files are projected to be operational beginning in January of 1991.

Open case files will contain DNA profiles of evidence samples (typically semen) left at the scenes of sexual assaults. The purpose of this file will be to link crimes committed by the same individual and thereby help investigators to coordinate their efforts to identify the perpetrator.

Missing Persons and Unidentified Deceased files could include DNA profiles of parents of missing persons as well as locate persons who are unidentified (that is, young children, Alzheimer's victims, deceased) and could serve to aid investigators and other authorities such as medical examiners in establishing identity.

The DNA indexes will be governed in large part by the progress made at the state level in acquiring the facilities and capability to perform DNA tests.

In order for these index files to be established, there must be agreement within the forensic community on testing standards which will provide the common language with which to store, collate and exchange DNA profile information. Other important benefits of standardization include improved quality control, easier training for personnel and the ability to respond to technology changes as a community so that we can adapt to new typing procedures more quickly. If we are working together and riding on the same train, as we change tracks we can do so in unison. As we need to generate new population frequency data with new probes, working together, we can do so in a fraction of the time it would take for us to build the data individually.

Jim Kearney has previously discussed the work of The Technical Working Group on DNA Analysis Methods (Kearney 1988). These individuals represent various crime laboratory systems and are addressing the technical requirements through a consensus building process. They have made great progress and have addressed the issues of standards, quality control guidelines and proficiency testing.

As Don Johnson (Johnson 1990) described for you, the National Crime Information Center (NCIC) computer system is an in-place, nationwide network accessible by police departments throughout the United States. We are examining this system as a repository or, at least, a communications network for DNA profile data.

The Advisory Policy Board provides guidance to the Director of the FBI for operating policies and guidelines for the NCIC system. The Advisory Policy Board consists of law enforcement representatives from all 50 states as well as representatives from selected interest groups. A DNA task force is being formed to examine the functional and operating requirements of DNA index systems and the feasibility of incorporating these systems with NCIC. The first meeting of the task force was held in September, 1989, and consisted of representatives of the Advisory Policy Board, The American Society of Crime Laboratory Directors, The National District Attorneys Association, and FBI policy-level managers from the NCIC and Forensic Science (DNA research programs).

Where are we now? Dr. Keith Monson (Monson and Budowle 1990) demonstrated the semi-automated analysis system to read and interpret the DNA autorads—the input device. This system is being installed in state and local crime laboratories that use the FBI DNA test protocol.

CONCLUSION

Woody Allen, the actor and comedian once said, "80% of success is just getting started." The use of DNA technology in law enforcement is off to a strong start. We are well on our way to seeing this powerful new technology effectively and responsibly applied. We are breaking new ground and must be open and flexible as we proceed. Thomas Jefferson wrote, "We are not afraid to follow truth wherever it may lead, nor to tolerate any error so long as reason is left free to combat it." And then let me leave you with the words of J. Edgar Hoover, an innovative leader who continuously stressed the need for communication, coordination, cooperation within the law enforcement community. He said, "The most effective weapon against crime is cooperation, the efforts of all law enforcement agencies with the support and understanding of the American people." Getting started, moving forward together with an open mind and spirit of cooperation and mutual support, we are bound to succeed in bringing the full potential from this powerful new technology to aid law enforcement and the criminal justice system in meeting the expectations of our citizenry.

REFERENCES

Adams, D. E., Presley, L. A., Deadman, H. A. and Lynch, A. G. (1990). DNA Analysis in the FBI Laboratory, In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C., pp. 173–177.

Baechtel, F. S. (1990). The Extraction, Purification and Quantification of DNA, In: Proceedings of

An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C., pp. 25–28.

Budowle, B. (1990). A Protocol for RFLP Analysis of Forensic Biospecimens. In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C., pp. 57–62

Budowle, B. and Monson, K. L. (1990). A Statistical Approach for VNTR Analysis. In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C., pp. 121–126.

Eubanks, W. G. (1990). Cost, Implementation and Training for DNA Analysis, In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C., pp. 189–193.

Jeffreys, A. J., Wilson, V., Wong, Z., Patel, I., Royle, N., Neumann, R., Armour, J. A. L., Kelley, R., Collick, A., Gray I. and Gibbs, M. (1990). Multilocus and Single Locus Minisatellite DNA Probes in Forensic Medicine, In: Proceedings of An International Symposium on the Foren-sic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C., pp. 77–86.

Johnson, D. M. (1990). The National Crime Information Center Past, Present and Future, In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C., pp. 133–135

Kearney, J. K. (1988). Summary of the first meeting of the Technical Working Group on DNA Analysis Methods, Crime Lab. Digest 15:115–117.

Monson, K. L. and Budowle, B. (1990). A System for Semi-automated Analysis of DNA Autora-diograms, In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C., pp. 127–132.

Parkin, B. (1990). DNA Analysis in the Metropolitan Police Forensic Science Laboratory, In: Proceedings of An International Symposium on the Forensic Aspects of DNA Analysis, U.S. Government Printing Office, Washington, D.C., pp. 163–167.

SECTION II

EXTENDED ABSTRACTS

BLANK PAGE

THE USE OF NON-ISOTOPIC DETECTION METHODS IN DNA FINGERPRINTING

R. Coquoz

Institut de Police Scientifique et de Crimonologie
Lausanne, Switzerland

The routine use of DNA fingerprinting would become more convenient and safer if the classical radioactive detection methods could be replaced by non-radioactive ones. Over the last few years there have been important developments related to the application of non-isotopic detection methods in molecular biology, evident by the appearance of several non-radioactive detection kits.

Several of these kits were examined in order to answer two questions; are any of these kits sensitive enough to detect a DNA fingerprint starting from a practicable amount of DNA? And which of these kits are the most sensitive? The kits tested were:

—The BlueGene kit (Gibco/BRL, Grand Island, NY) based on a nick-translation using biotin-11-duTP, and detection with a streptavidin-alkaline phosphatase conjugate

—The Boehringer kit (Boehringer, Mannheim, W. Germany) based on a random primer labelling system using digoxigenin-dUTP, and a detection with an anti-digoxigenin antibody-alkaline phosphatase conjugate

—The ChemiProbe kit (FMC Bioproducts, Rockland, ME) based on the chemical modification of the DNA probe, and a detection with an anti-modified DNA antibody-alkaline phosphatase conjugate

—The BioBridge kit (ENZO Biochem, New York, NY) based on the end-labelling of oligonucleotide probes to get a polyT tail, and a detection with biotinylated polyA and streptavidin-alkaline phosphatase complex.

These kits were chosen as they are based on different principles. The same type of kits from various suppliers (for example different streptavidin-alkaline phosphatase suppliers) were not compared, because major differences were not expected.

The Southern blots were made from 20 cm long, 0.6% agarose gels run for 60 hr. at 50 V in JAE buffer which were blotted onto nitrocellulose BA85 (Schleicher & Schuell, Feldbach, Switzerland) using the high salt capillary blotting protocol proposed by the supplier. The DNA was fixed at 80° C for the 2 hr. using an IR lamp. All the hybridizations have been done in 5 x SSC at 42° C for 24 hr. 0.02% SDS, 0.1% N-Lauroylsarcosin Nasalt, 5% Boehringer blocking mixture and 15% formamide. A prehybridization was done in the same solution for 8 hr. The membrane were washed at 50° C in 2 x SSC, 0.1% SDS. The probe used with the BioBridge kit was an oligonucleotide, 48 nucleotide long, corresponding to 3 repeats of the lambda 33.15 probe (Jeffreys et al. 1985). The EcoRI linearized M13 phage (Vassart et al. 1987) was used with the 3 other kits.

Each kit was used according to its instructions. The controls provided with the kits were tested to check our ability to reach the performances claimed by the supplier. In each case, we reached the subpicogram range promised. Without a direct comparison possible, we could only assume that our sensitivity was equivalent or close to the optimal.

Figure 1 answers the first question positively. It is indeed possible to detect a satisfactory DNA fingerprint from less than 10 μg DNA (or about 300 μl of blood). This sensitivity is not sufficient to satisfy the needs of routine forensic laboratories which will still have to rely on P^{32} with a sensitivity down to 0.2 μg (Bär et al. 1989). However, our results give us hope that developments in these techniques might bring, in the very near future, a true alternative to radioactive detection.

Among the various kits tested, the best results were obtained with the ChemiProbe and the Boehringer kit.

REFERENCES

Bär, W., Kratzer, A., Malchler, M. and Schmid, W. (1989). Postmortem stability DNA, Forens. Sci. Int. 39:59–70.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985). Hypervariable minisatellite regions in human DNA, Nature 314:67–73.

Vassart, G., Georges, M., Monsieur, R., Brocas, H., Lequarre, A. S. and Christophe, D. (1987). A sequence in M13 phase detects hypervariable minisatellite in human and animal DNA, Science 235:683–684.
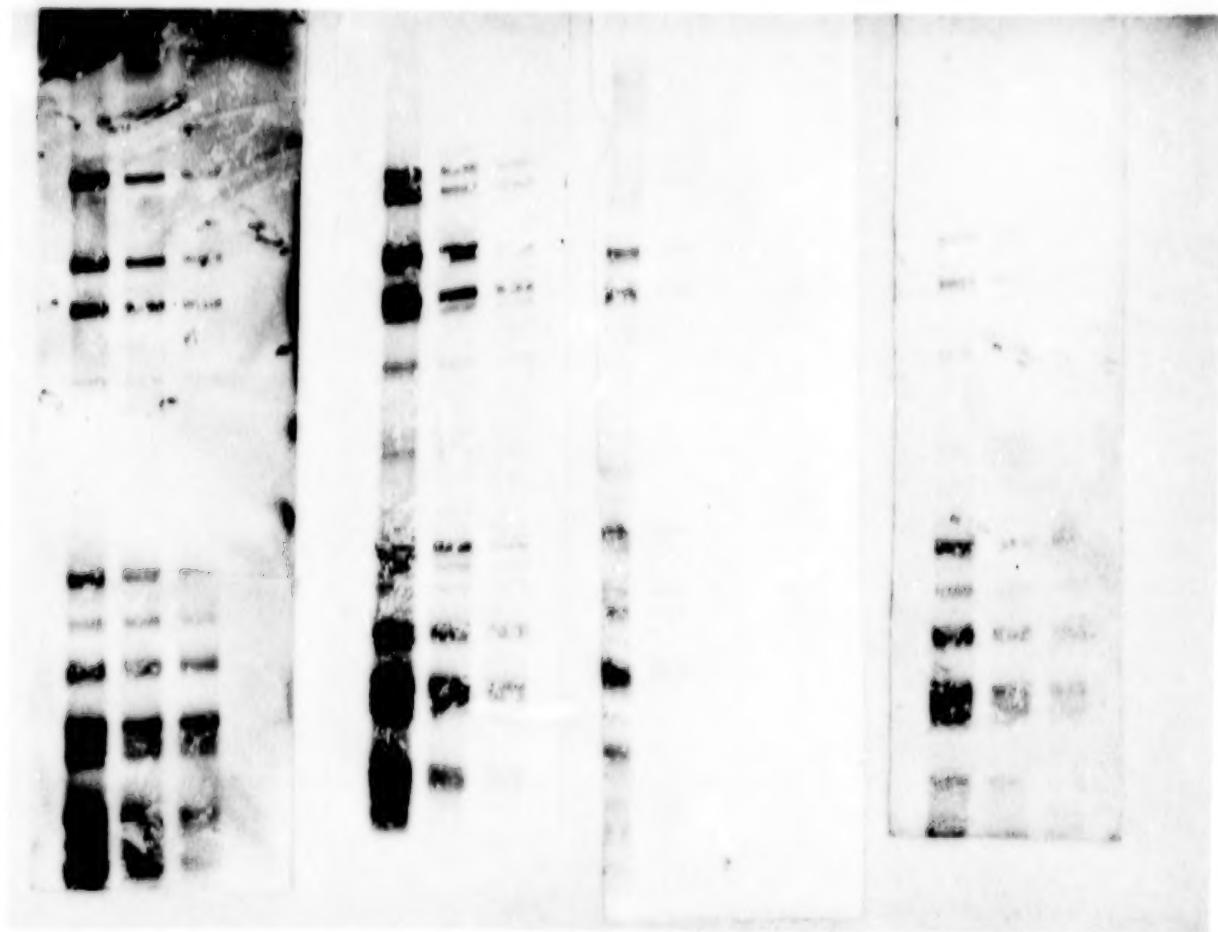
Figure 1. Southern blots each containing 3 lanes with 100, 25 and 100 μg respectively of *Hae*III digested human genomic DNA. From left to right, the blots were processed using the Boehringer kit, the Chemiprobe kit, the BioBridge kit and BlueGene kit.

THE REPRODUCIBILITY OF CRI PROBE L336 FOR IDENTIFICATION AND PARENTAGE ANALYSIS

P. J. Lincoln, C. P. Phillips, J. A. Thomson, P. H. Watts and N. J. Wood

London Hospital Medical College
London, England

The practical application of Collaborative Research Incorporated probe L336 (Donis-Keller *et al.* 1987) for identifying individuals and for assessing the probability of paternity was attempted in a small series of blood samples from families requesting paternity investigations. The reproducibility of the results obtained from the use of L336 is of concern, because this probe detects a large number of alleles or variants. Many of these are relatively closely spaced in molecular size.

The reproducibility was investigated by the molecular sizes of the DNA fragments containing L336 sequences in individuals when their DNA was analyzed on separate gels and by estimating the molecular sizes of the L336 sequence-containing fragments in parent/child pairs on the same gel.

Maternity was assumed, and the biological relatedness of the father was ascertained using a range of conventional blood group and serum or blood cell enzyme polymorphisms producing a high level of correctly ascribed paternity (with Paternity Indices >100) probability. DNA was prepared as described by Gill *et al.* (1987) and digested with *Pst*I following the manufacturer's directions (Northumbria Biologicals Limited, Cramlington, UK). The digested DNA was examined by electrophoresis using a small sample on a 0.8% Agarose minigel in TBE. If the DNA appeared to be completely digested, it was then subjected to electrophoresis on a 25 x 20 cm 0.85% Agarose gel in TAE at 30V, 40 mA for 45 hrs. until fragments of 1.5 Kb had reached the end of the gel. The gel was then soaked in 0.4 M NaOH, 1.5 M NaCl for 30 min., then neutralized in 0.5 M Tris Cl, 1.5 M NaCl pH 7.2, and the DNA was Southern transferred onto a Hybond N membrane (Amersham International, Amersham, UK) with 10 x SSC. Five hundred ng of Plasmid containing probe L336 was radiolabelled with P^{32}–dCTP using the multiprime method (Feinberg and Volgelstein 1983) and hybridized with the membrane-bound DNA in CRI hybridization solution at 50° C for 18 hrs. After hybridization the membrane was washed briefly in 2 x SSC and then twice in 0.1 x SSC, 0.2% SDS at 65° C for 45 min. as recommended by CRI. Autoradiographs were prepared exposing the film to the membrane with two intensifying screens for 10-14 days at −80° C.

The molecular sizes of the DNA fragments containing the L336 sequences were estimated from their migration distances by comparison with those of molecular size markers on the same gel. Molecular size markers of 9.69, 7.74, 6.22, 4.25, 3.47, 2.69, 2.39 and 1.88 Kb were run in three of the lanes in each gel. The log 10 marker size was plotted against the migration distance for each of the 3 lanes and straight lines fitted through the plots by the least-squares method. The molecular sizes of the L336 bands were then estimated by measuring their electrophoretic migration distances and interpolating them on the plot of marker size against distance.

An average variation of 1.9% (without a maximum of 7.2%) was seen when repeated analyses were made on separate DNA gels from the same individual samples. When estimates of the molecular sizes of L336 containing fragments in parent/child pairs were made, having analyzed their samples simultaneously on the same gel, an average variation of .7% (range 0 to 4%) was seen. See Table 1.

Table 1. VARIATION IN ESTIMATED SIZE OF DNA BAND

(a) when the band is seen in a parent and his/her child
(b) when the same individual is tested on separate occasions.

	number of pairs	mean percent difference in size	maximum seen
(a) Adjacent lanes on the same gel:			
mother and child	47	.72	4.0%
father and child	28	.65	4.1%
(parent and child	75	.68)	
(b) Separate gels	74	1.88	7.2%

These figures indicate the level of caution required in deciding the identity of L336 bands when comparing DNA from different sources.

We gratefully acknowledge the advice and encouragement of Collaborative Research Incorporated, and their generous gifts of the probe.

REFERENCES

Donis-Keller et al. (1987). A genetic linkage map of the human genome. Cell 51:319–337.

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity, Anal. Biochem. 132:6–13.

Gill, P., Lygo, J. E., Fowler, S. J. and Werrett, D. J. (1987). An evaluation of DNA fingerprinting for forensic purposes, Electrophoresis 8:38–44.

APPLICATION OF DNA FINGERPRINT ANALYSIS WITH MINISATELLITE DNA PROBES TO INDIVIDUAL IDENTIFICATION

H. Akasaka and S. Nonomura

Metropolitan Police Department
Tokyo, Japan

We used this method in order to apply the fingerprint analysis described by Jeffreys *et al.* (1985) to individual identification for forensic science. We initiated an analysis of sample DNA by Southern blot hybridization using two minisatellite DNAs and variable number tandem repeat (VNTR) clones as probes. Further, we attempted to amplify sample DNA of a VNTR clone's locus using the polymerase chain reaction (PCR) method. We ultimately subjected the sample to gel analysis.

MATERIALS AND METHODS

High molecular weight DNA was isolated from human lymphocytes of 20 healthy adults who were not related consanguineously. As specimens, DNA samples were isolated from human sublingual glands, kidneys and blood. The DNA fragments digested with restriction enzyme *Hae*III or *Hinf*I were subjected to electrophoresis on 0.7 or 1.0% agarose gel for 15–18 hr. After transferring the DNA fragments to a nylon filter, each DNA on the filter was hybridized with P^{32}-labeled minisatellite probes such as core sequence or M13 phage DNA and a VNTR clone, pYNH24, isolated by Nakamura *et al.* (1987). Autoradiography occurred at $-85°$ C for 15–39 hr. As for 3 DNA specimens, the PCR method described by Saiki *et al.* (1988), using the 2 primers reported by Horn *et al.* (1989), amplified the VNTR locus of pYNZ22.

RESULTS

With a minisatellite core and M13 phage DNA probes, we obtained highly polymorphic autoradiographic patterns and among 10 people, no same band-patterns appeared. The same DNA samples showed completely different patterns with the 2 different minisatellite probes.

With a VNTR clone of pYNH24, 2 allelic bands appeared distinctively among another set of DNA samples from 10 people. Then we applied this probe to an examination of individual identifi-

cation or parent-child relationships. In one case, two bands detected by pYNH24 were identical between DNAs from the upper and lower halves of the body. This detection suggests that the halves were derived from the same body. In another case, we attempted to determine parent-child relationships using the pYNH24 and pYNZ22 probes. A suspected woman's DNA, after pYNH24 probing, showed the identical 2 bands of her parents. The same result occurred with sample DNAs amplified at the pYNZ22 locus using the PCR. Thus the results with two probes suggested that the woman is really the parents' child.

Then we showed that minisatellite or VNTR DNA clones are useful to identify individuals. Amplification at a certain VNTR locus of PCR seemed to be an especially timesaving and informative method, as was nonisotopical detection.

REFERENCES

Horn, G. T., Richards, B. and Klinger, K. W. (1989). Amplification of a highly polymorphic VNTR segment by the polymerase chain reaction, Nucl. Acids Res. 17:2140.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985). Hypervariable 'minisatellite' regions in human DNA, Nature, 314:67–73.

Nakamura, Y., Leppert, M., O'Connel, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping, Science, 235:1616–1622.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, Science, 239:487–491.

THE USE OF POLYSTAT 3 PROBE FOR FORENSIC TESTING.

W. J. Harding, A. M. Ross and J. C. S. Fowler

Forensic Science Centre
Adelaide, South Australia

J. L. McInnes

University of Adelaide
Adelaide, South Australia
Bresatec Limited
Adelaide, South Australia

Polymorphism in repetitive DNA is generally associated with tandemly repetitive sequences and results from variation in the number of tandem repeats at any particular genomic site, hence the description variable number tandem repeats (VNTR). Two different classes of VNTR variation can be defined based upon DNA sequence enzyme restriction sites (Fowler *et al.* 1988a). Restriction may be in sequences flanking the VNTR repeats, such examples being commonly called mini-satellites. Alternatively, restriction may occur within the long tandem arrays at point mutatable sites which are distributed unevenly throughout the sequence. The polymorphism detected using the probe Polysat 3™ (Bresatec, Adelaide, South Australia) on *Taq*1 digests of human DNA is an example of this second type. The polymorphism probably arises from C to G point mutations in the pentameric repeat sequences TTCCA which are common in Satellite III (Fowler *et al.* 1988b).

Details of DNA sample preparations have been published (Fowler *et al.* 1987) as have the field inversion electrophoretic conditions (Fowler *et al.* 1988c). Following electrophoresis samples are blotted to nitro-cellulose membranes as described by McInnes *et al.* (1987). Probes are biotinylated using Photobiotin™ according to the method of Forester *et al.* (1985), and the bands developed using an avidin-alkaline phosphatase conjugate.

With *Taq*1 digests of human DNA the Polysat 3™ probe produces complex, multi-banded RFLP patterns which are highly discriminating for individuals (Fowler *et al.* 1988b). The bands range in size from about 5 to 25 Kb. Improved resolution of the higher molecular weight bands (about 9 to 20 Kb) achieved using field inversion gel electrophoresis on 1% agarose gels makes the patterns easier to compare and readily amenable to automated image capture and processing. In contrast to Polysat 3™, Probe 216s (another probe also derived from sperm) gives an invariant pattern for all individuals. It can be used in conjunction with Polysat 3™ as a quality control procedure to check both the quantity of DNA and quality of the restriction digestion.

With *Hae*III digests Polysat 3™ detects a 3.4 Kb fragment in male DNA only (Y-chromosome) with Probe 216s again giving an invariant pattern. Since the Probe 216S bands are clear of the 3.4 Kb gel region, both probes can be used as a mixture to provide the quality control necessary to ensure that both male and female DNA are correctly determined.

The high information content of the multi-banded patterns produced by Polysat 3™ with *Taq*1 digests, coupled with the quality control available using Probe 216S makes this combination a very useful tool in family studies and parentage testing. Although the photobiotin labelling has been found to be less sensitive than radioactive probes, the non-isotopic method is cheap, convenient and safe, as well as more rapid.

(Polysat 3 and Photobiotin are registered Trade Marks of Bresatec, Ltd.)

REFERENCES

Forster, A. C., McInnes, J. L., Skingle, D. C. and Symons, R. H. (1985). Non-radioactive hybridization probes prepared by the chemical labelling of DNA and RNA with a novel reagent, photobiotin, Nucleic Acids Res. 13:745–761.

Fowler, J. C. S., Harding, H. W. J. and Burgoyne, L. A. (1987). A protocol using an alkali blotting procedure for the analysis of restriction length fragments of human DNA, Adv. Forensic Haemogentics 2:339–346.

Fowler, J. C. S., Burgoyne, L. A., Harding, H. W. J. and Scott, A. C. (1988a). Repetitive deoxyribonucleic acid (DNA) and human genome

variation—A concise review relevant to forensic biology, J. Forensic Sci. 33:1111–1126.

Fowler, J. C. S., Drinkwater, R., Skinner, J. D. and Burgoyne, L. A. (1988b). Human satellite III DNA: an example of a macrosatellite polymorphism, Hum. Genet. 79:265–272.

Fowler, J. C. S., Skinner, J. D., Burgoyne, L. A. and McInnes, J. L. (1988c). Improved separation of multi-locus hypervariable DNA restriction fragments by field inversion gel electrophoresis and fragment detection using a biotinylated probe, Appl. and Theoret. Electrophoresis 1:23–28.

McInnes, J. L., Dalton, S., Vise, P. D. and Robins, A. J. (1987). Non-radioactive photobiotin—labelled probes detect single copy genes and low abundance mRNA, Biotechnology 5:269–272.

214

IDENTIFICATION OF GENOMIC DIGESTS FROM SEVERAL SOMATIC SOURCES BY MULTI-LOCUS AND LOCUS-SPECIFIC DNA PROFILES. A SURVEY OF SOME CRIMINAL CASES WORKED OUT IN THE UCSC AND SPS LABORATORIES

V. L. Pascali, E. d'Aloja and M. Dobosz

Universita Catolica del S. Cuore

Rome, Italy

A. Spinella and A. Quaresima

Servizio di Polizia Scientifica

Rome, Italy

Fourteen criminal cases involving DNA profile analysis of various biological specimens have been worked out recently by our molecular biology laboratory unit.

On such occasions, analyses of differently aged genomic DNA from various somatic sources (blood stains, hair roots, chorionic villi and fetal tissues) were performed. Here we refer to our results, with a view to focus on technical problems we encountered, and to draw a statistical briefing of our yet limited experience.

MATERIALS AND METHODS

Standard Southern blot analysis was applied to *Hin*fI digests of every specimen to be analysed, according to Maniatis *et al.* (1982). Modifications to the general procedure described therein were introduced at the stages of DNA extractions (Gill *et al.* 1985) and filter hybridizations.

Filter blots (Hybond N) were hybridized to alpha globin 3'HVR probe under either relaxed and/or stringent conditions. Probes YNH24 (Nakamura *et al.* 1987) and lambda MS1/lambdaMS31 (Wong *et al.* 1987) were also used.

RESULTS

The most outstanding features of the biological specimens we analyzed are summarized in Table 1, together with the hybridization procedure(s), and their outcome. Over 60% of the specimens yielded a coherent band pattern. Some pitfalls are ascribed to analyses of hair roots, none of which succeeded

in giving DNA bands. We ascribe this result to the fact that a few hairs could be pooled together in each case (following simple morphological criteria) and to an unexpectedly fast aging of roots cells. Consequently, we presently disregard Southern blot analysis for hair identification analysis. Additional drawbacks were met with vaginal/sperm cells and blood stains, due to the relatively poor content in high-MW DNA in aged/degraded samples. However, unlike hair samples, degraded stains were at times typed by rehybridizing blots to supplementary probes. This suggests a somewhat whimsical behavior of DNA degradation, where detrimental effects can be, at times, circumvented by multiple probing.

REFERENCES

Gill, P., Jeffreys, A. J. and Werrett, D. J. (1985). Forensic applications of DNA fingerprints, Nature 318:577–579.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). Molecular cloning, Cold Spring Harbor Laboratory Press.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumplin, E. and White, R. (1987). Variable number of tandem repeats (VNTR) markers for human gene mapping, Science 235:1616–1622.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. J. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA, Ann.Hum.Genet. 51:269–288.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. J. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA, Ann.Hum.Genet. 51:269–288.

TABLE 1

	Tissue	Substrate/Origin	Age	Approximate Quantity	DNA MW	Probe/String	Bands Det.	Outcome
Case 1	blood	cloth	24	abundant	H	3'HVH/lo	10	excl.
	blood	gravel		abundant	D	3'HVR/lo	none	
Case 2	blood	cloth	7	medium	H	3'HVR/lo	13	excl.
Case 3	blood	cloth	1	abundant	H	3'HVR/lo	20	ident.
	blood	linen	1	abundant	H	3'HVR/lo	20	ident.
Case 4	villi	/	2	abundant	H	YNH24/hi	2	excl.
	fetal	spleen	2	abundant	H	YNH24/hi	2	excl.
		lung		abundant	H	YNH24/hi	2	excl.
		thymus		abundant	H	YNH24/hi	2	excl.
		heart		abundant	H	YNH24/hi	2	excl.
Case 5	blood	cloth	5	scarce	D	3'HVR/hi	none	
	blood	cloth	5	abundant	H	YNH24/hi	2	
Case 6	hairs	slide-mounted	18	10	/	3'HVR/hi	none	
						YNH24/hi	none	
Case 7	blood	cloth	7	abundant	D	3'HVR/hi	none	
					D	YNH24/hi	none	
	blood	cadaveric		medium	D	3'HVR/lo	14	
Case 8	blood	car seat	7	scarce	H	3'HVR/hi	2	excl.
		car seat		scarce	H	3'HVR/hi	2	excl.
		cloth		abundant	H	3'HVR/hi	2	ident.
Case 9	blood	cloth	4	scarce	/	3'HVR/lo	none	
						lambda MS1/31	none	
Case 10	blood	ground	2	scarce	D	3'HVR/hi	none	
						lambda MS1/31	none	
Case 11	womb	cadaveric	6	abundant	H/D	3'HVR/hi	2	
	sperm	cadaveric (same as above)		medium	D	YNH24/hi 3'HVR/hi	none none	
Case 12	sperm	condom	7	abundant	D	3'HVR/hi YNH24/hi lambda MS1/31	none none	
Case 13	sperm	ground	1	medium	D	3'HVR/hi	none	
						YNH24/hi	none	
Case 14	blood	paper towel	6	abundant	H	YNH24/hi	2	ident.

Legend to Table 1: H: high molecular weight DNA; D: degraded DNA; hi: high stringency; lo: low stringency; ident: identification; excl.: exclusion; terms as "abundant", "medium" and "scarce" pertain to a rough, first-sight evaluation of samples. Age of samples expressed in months.

DNA PROFILES FROM DENTAL PULP

M. E. Lawton, P. Stringer and M. Churton

Chemistry Division
Auckland, New Zealand

The introduction of DNA profiling into Forensic Serology has dramatically increased the discriminating power of blood or body fluid characterization.

Applications other than in crime investigations are also obvious. The technique lends itself very well to parentage testing and could be used in cases of mass disasters or possibly for identifying an unidentified corpse. In those latter cases the bodies may be badly mutilated, burned or degraded. Blood may not be available, and another source of DNA will need to be obtained. Any other nucleated cell can potentially be used to provide the necessary starting material. One source of DNA which is relatively well protected from mechanical damage is dental pulp. Because of the inherited nature of DNA polymorphisms, DNA extracted from the dental pulp of an unknown corpse could be compared with DNA profiles from the blood samples of possible close relatives.

Dental pulp is a connective tissue which is mainly comprised of collagen fibers and cellular matter. In young pulp, collagen fibers are found around the blood vessels as supporting elements. As the individual ages, more collagen is deposited, resulting in a harder dental pulp.

MATERIALS AND METHODS

Undecayed wisdom teeth were used for this study with corresponding NaEDTA blood samples where appropriate. The donors ranged in age from 16 to 86 years old. Some teeth were stored at $-4°$ C for a short time before DNA extraction, while others were subjected to a detrimental environment for several weeks.

The teeth were broken mechanically, and the dental pulp was removed. Dental pulp and whole blood were phenol extracted according to the method of Jeffreys *et al.* (1985). All samples were checked for high molecular weight content on a minigel. Some corresponding blood and teeth samples were digested overnight with *Hinfl*, blotted onto nylon membrane and probed with P^{32}labelled 33.15.

RESULTS

Dental pulp was shown to be a concentrated source of DNA. While the amount of DNA extracted was as expected generally less for older persons, this was not always a linear correlation. Samples from a person in his twenties had no extractable DNA while the oldest person to yield sifficient DNA for profiling from a single tooth was 48 years.

In all cases, even freshly extracted teeth, the DNA from dental pulp was to some degree more degraded than the fresh or preserved corresponding blood sample. They did however contain sufficient high molecular weight DNA to obtain a final profile although the effects of degradation could sometimes be observed (Figure 1).

The quantity and quality of DNA from teeth which had been buried for several weeks showed only minor deterioration compared with well preserved teeth from the same persons. Prolonged storage of teeth in high heat and humidity produced shrinkage of the dental pulp and had caused it to disappear altogether in some cases.

Dental pulp from teeth can provide a source of DNA for profiling in cases of unidentified bodies and some criminal instances. The amount of available DNA varies considerably from one person to another and is to some extent age dependent. Conditions of high humidity will cause degradation of dental pulp thus reducing the chance of obtaining a profile.

REFERENCE

Jeffreys, A. J., Wilson, V. and Thein, S. L., (1985). Nature 314:67–73.
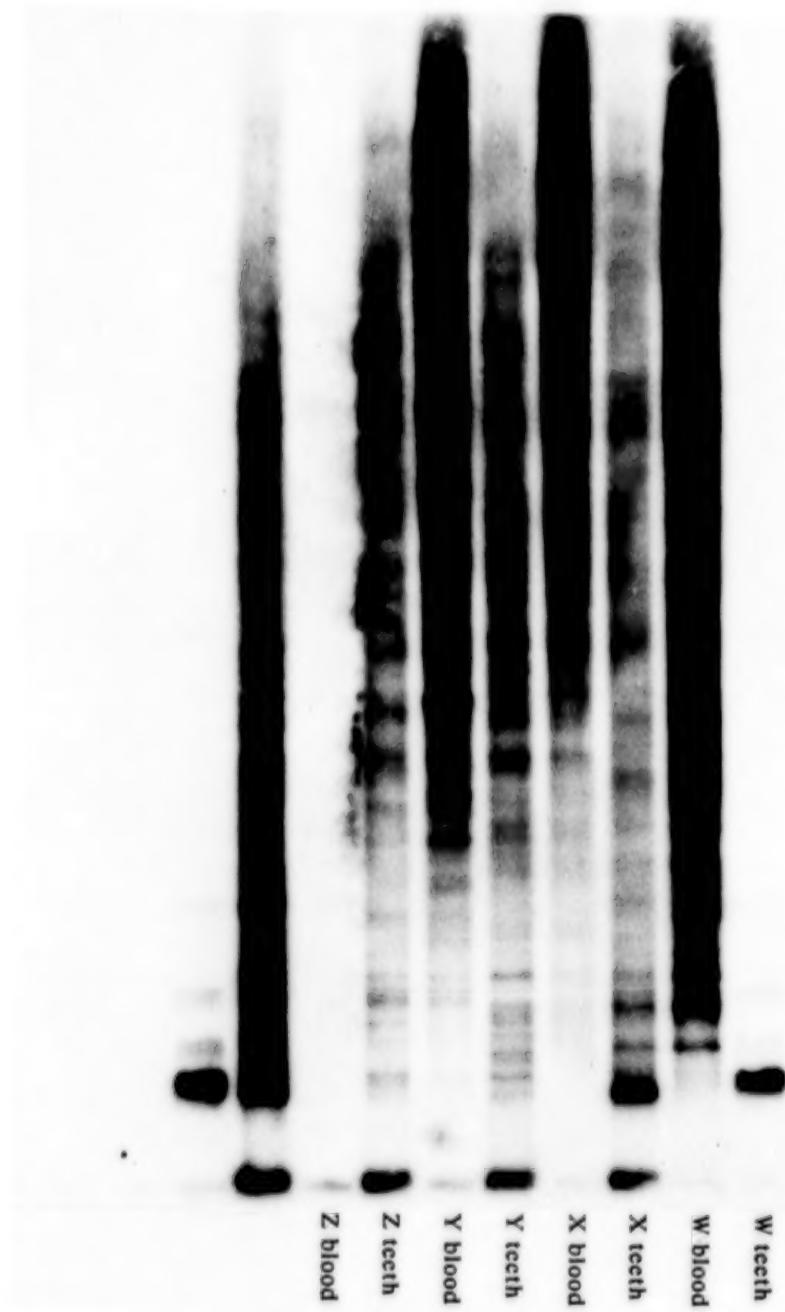
Figure 1. DNA profiles from teeth and corresponding blood samples

RESTRICTED FRAGMENT LENGTH POLYMORPHISM (RFLP) ANALYSIS OF BIOLOGICAL STAIN MIXTURES

L. A. Presley and D. E. Adams

FBI Laboratory
Washington, D.C.

Biological fluid evidence in criminal cases is often contaminated and/or mixed prior to submission to the crime laboratory for DNA analysis; therefore, it is of paramount importance that the impact of environmental factors and body fluid mixtures be assessed. Several experiments were designed to determine the effects of chemical and bacterial contamination, environmental exposures, and biological fluid mixtures on the recovery, quantity and quality of DNA in laboratory prepared body fluid stains. The FBI Laboratory restriction fragment length polymorphism (RFLP) protocol for DNA analysis (Budowle 1988) was used to analyze the samples.

CHEMICAL CONTAMINATION

A variety of commonly encountered chemical substances which included unleaded gasoline, motor oil, detergent, acid, base, salt, bleach and soil were mixed with blood samples. High molecular weight DNA was recovered after 5 days, and RFLP banding patters consistent with the uncontaminated bloodstain from the same origin were obtained from all the contaminated samples except for soil (Figure 1).
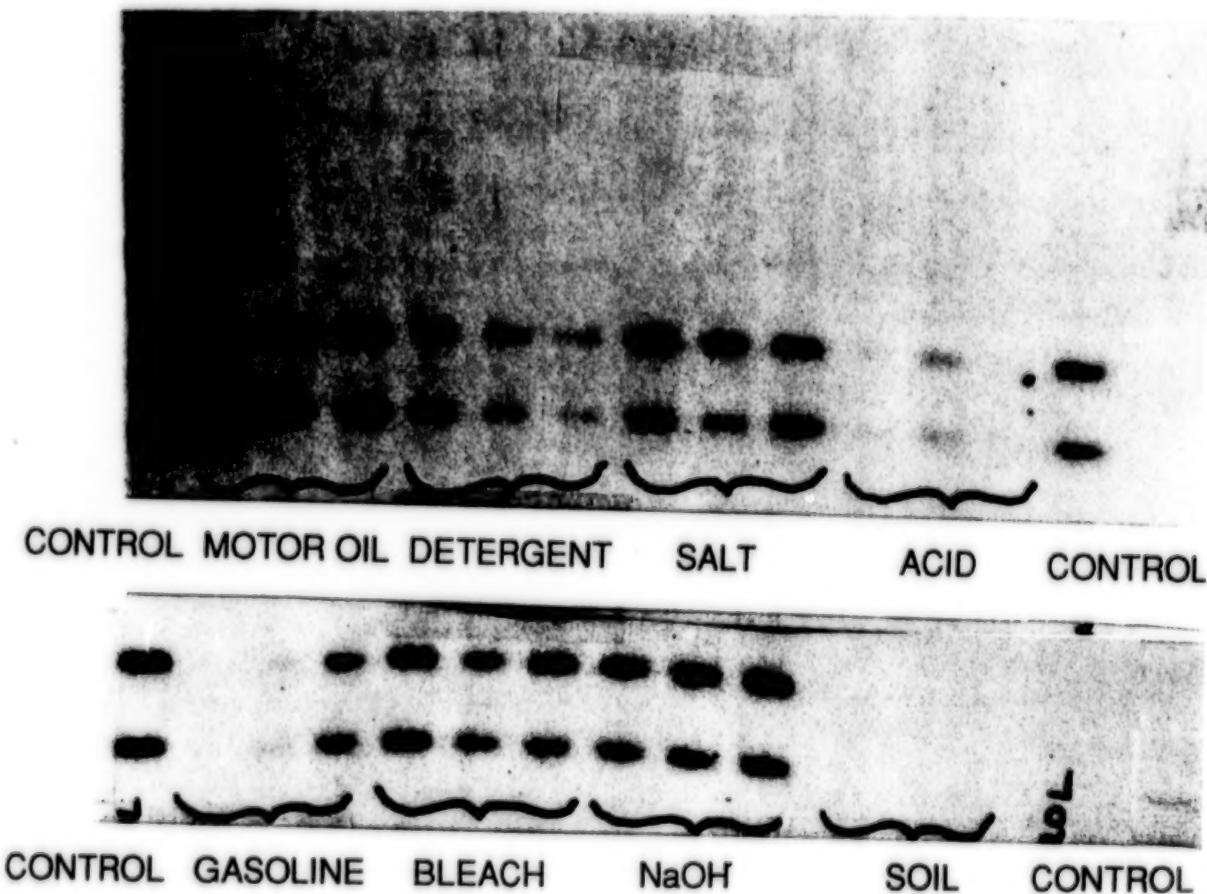
Figure 1

BACTERIAL CONTAMINATION

Microorganisms which included *Staphloccus epidermis, Candida valida (yeast), Escheria coli* and *Bacillus subtilis* were mixed with blood or semen, made into stains and sampled in 5 days. High molecular weight DNA recovered from the stains, and the RFLP banding patterns were consistent with the uncontaminated samples (Figure 2). No hybridization occurred with the DNA from the microorganisms.

ENVIRONMENTAL EXPOSURES

Bloodstains of known origin were exposed to ambient outdoor temperatures for approximately 3 months. After 8 weeks of exposure to sunlight, the DNA extracted from the bloodstains was degraded to the point that no RFLP banding patterns were detected. In a similar experiment, bloodstains were exposed to sunlight and were harvested daily. Stains in this study that were maintained in the sunlight for 10 days resulted in extracted DNA which was degraded, while duplicate samples stored at the same temperature in the dark produced acceptable RFLP banding patterns (Figure 3).

In a separate experiment, bloodstains were prepared on a variety of substrates which included cotton, nylon, denim, glass, wood and aluminum. These bloodstains were then subjected to 1 of 4 different temperatures: 4°, 22°, 37° C and ambient temperature. After 2 days, 1, 3 and 5 month periods of exposure, with minor exceptions, all the samples on all the substrates at all temperatures produced RFLP banding patterns consistent with the untreated known liquid blood samples (Figure 4).

BIOLOGICAL FLUID MIXTURES

An array of body fluids from separate donors including semen, blood, urine, saliva and vaginal secretions were mixed. The RFLP banding patterns were consistent with the known donor samples obtained from mixtures containing blood, semen and vaginal secretions, except in the mixtures containing only 1µl of the body fluid. Approximately 10µl samples of saliva were rarely detected and no RFLP banding patterns for urine were detected (Figure 5).

CONCLUSIONS

These studies involved thousands of samples and resulted in autoradiographs which were either consistent with the known samples or no results were obtained due to the predictable influences on the integrity of the DNA. These results strongly support the FBI protocol for DNA analysis by the RFLP technique and demonstrate the reliability and reproducibility of the technique under a variety of commonly encountered conditions.

REFERENCE

Budowle, B. (1988). The RFLP Technique, Crime Lab. Digest, 15:97–98.

CONTROLS BACILLUS SUBTILIS   BLOOD + BACTERIA   SEMEN + BACTERIA

Figure 3

Figure 4

221

BLOOD/VAGINAL   SALIVA/VAGINAL CONTROL URINE/VAGINAL   SEMEN/VAGINAL   SEMEN/VAGINAL   MALE FEM   SEM/VAG
10 ul blood   10 ul saliva   10 ul urine   10 ul semen   10 ul semen   BLOOD   MALE / FEM

Figure 5

222
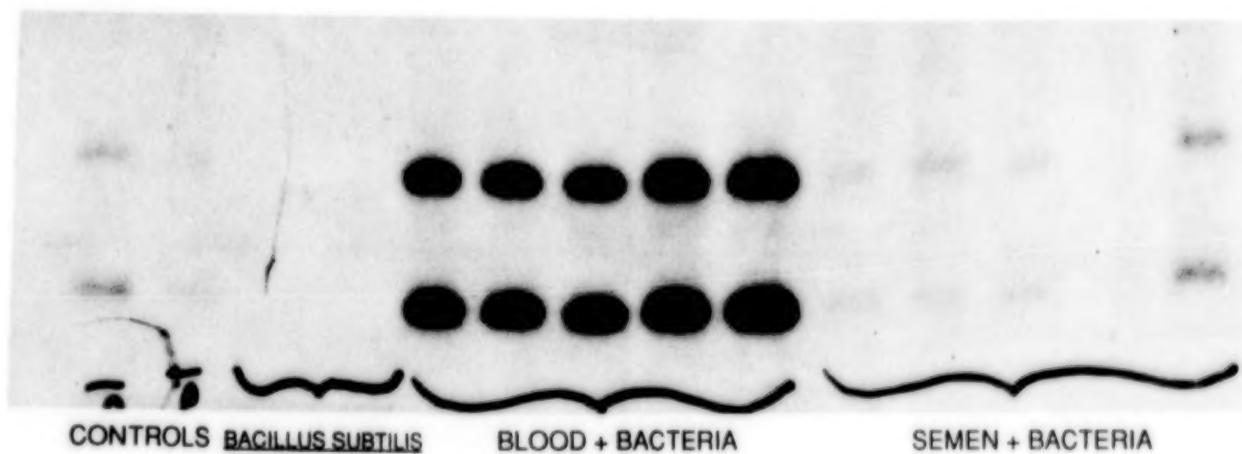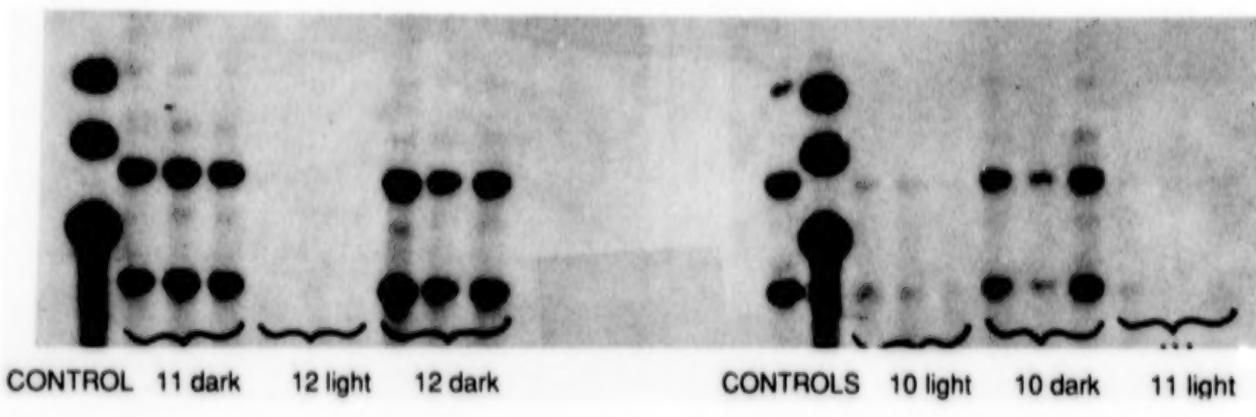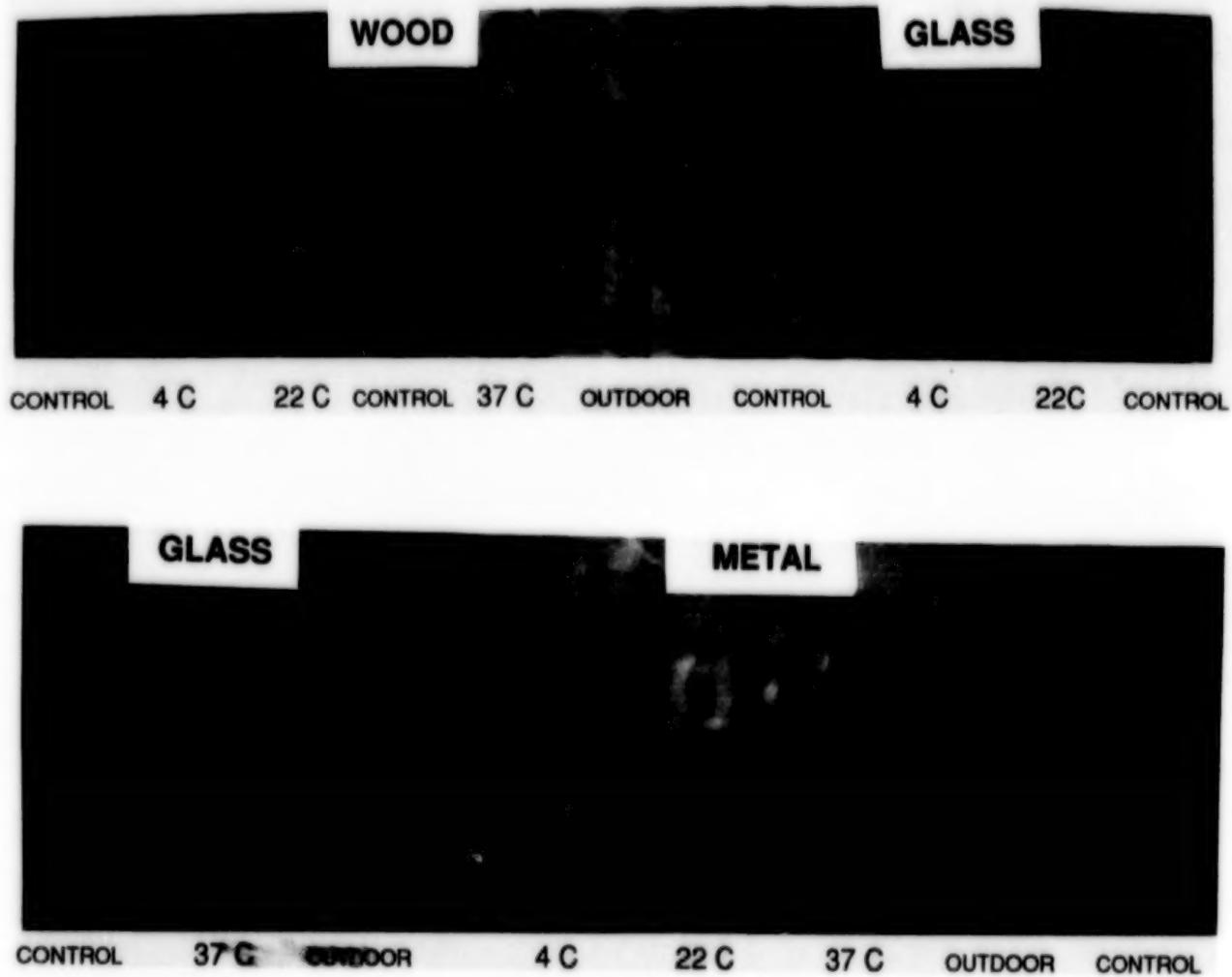
HUMAN IDENTIFICATION WITH SYNTHETIC OLIGONUCLEOTIDE PROBES

B. W. Kovacs, B. Shahbahrami and D. E. Comings

City of Hope Medical Center
Duarte, California

B. L. Johnson

Los Angeles County Sheriff's Department
Los Angeles, California

The human genome contains many hypervariable regions (HVRs) consisting of short core sequence elements repeated in tandem arrays (minisatellites). The exceptional polymorphic variability occurring at these loci is due to frequent variation in the number of tandem repeat (VNTR) elements present. Due to moderate sequence homology, low stringency hybridization of cloned DNA probes containing segments of core repeat elements simultaneously detect multiple hypervariable loci. This has been exploited to produce individual specific hybridization patterns (DNA fingerprints). However, only moderate sequence homology introduces uncertainty in determining loci detected (Ali and Wallace 1988).

The thrust of these studies has been to improve this situation by designing and characterizing new molecular probes for human identification. We have focused on using synthetic oligonucleotide probes composed of trinucleotide core repeats which makes this technology simpler, more accurate and more readily available to law enforcement agencies.

MATERIALS AND METHODS

Synthetic oligonucleotide probes (24 mers) composed of simple and mixed trinucleotide repeat core sequences were produced on an automated DNA synthesizer. Purified probes were radiolabeled with P^{32} by T4 kinase. Hybridization was carried out in gel (Ali and Wallace 1988) on 3-7 μg restriction digests (*Hinf*I, *Hae*III, *Sau*3A, & *Taq*I. BRL, Bethesda, MD) of human DNA after electrophoresis at 45 V for 40 hr. High stringency was obtained using tetramethylammonium chloride post hybridization washes as described by Wood et al. (1986). Autoradiographs were obtained after 24 hr. at 70° C without intensifying screens. Estimates of band sharing frequency were obtained by pairwise comparisons of 30 random individuals as described by Jeffreys *et al.* (1985a). Family studies were done using the C.E.P.H. family panel.

RESULTS

Most of the oligonucleotide core repeat probes identified complimentary sequences at high stringency. Many produced hybridization patterns of similar complexity to those obtained with Minisatellite probes of Jeffreys *et al.* (1985b). Table 1 summarizes the characteristics of the loci detected by each probe. The complex patterns were obtained using several different endonucleases, indicating VNTR type polymorphism described by Nakamura *et al.* (1987). Family studies demonstrated that the fragments detected were inherited in Mendelian fashion and correspond to multiple unlinked loci. Several of the probes were used to screen unlinked loci. Several of the probes were used to screen human cDNA libraries at high stringency. A large number (> 300) of positive clones were isolated (Kovacs *et al.* 1988), indicating that several of these new hypervariable loci are expressed in humans.

Table 1. POPULATION CHARACTERISTICS OF LOCI DETECTED

Repeat Seq.	Frags/Indiv.	Unique/Indiv.	Band sharing
CCA/GGT	24.9	17.1	0.22
CCG/GGC	16.5	8.3	0.51
AAC/TTG	17.4	4.3	0.81
TCT/CCT	10.3	5.1	0.86
CAG/CCA	26.8	19.2	0.14
AAG/TTC	12.6	8.6	0.33

CONCLUSIONS

These studies indicate that short oligonucleotide probes composed of trinucleotide core repeats detect many new hypervariable VNTR type loci. Further, because short oligonucleotides are highly sequence specific at our conditions, these loci share exact core sequence homology. This improves accuracy, because a single base mismatch will prevent hybridization. The probes detecting the larg-

est numbers of VNTR loci each produce different individual specific molecular genetic fingerprint patterns. Thus these oligonucleotide probes enlarge the repertoire of loci useful in human identifications. In addition, the use of synthetic probes lends itself to rapid in gel hybridization and obviates cloning. Work is in progress to characterize the cDNA clones identified.

REFERENCES

Ali, S. and Wallace, R. B. (1988). Intrinsic polymorphism of variable number tandem repeat loci in the human genome, Nuc. Acid Res. 16:8487–8496.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985a). Individual-specific fingerprints of human DNA, Nature 316:76–79.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985b). Hypervariable minisatellite regions in human DNA, Nature 314:67–73.

Kovacs, B. W., Shahbahrami, B. and Comings, D. E. (1988). Polymorphic variation and dispersion of codon tandem repeats in the human genome, Am. J. Hum. Genet. 43:A757.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping, Science 235:1616–1622.

Wood, W.I., Gitschier, J., Lasky, L. A. and Lawn R. M. (1985). Base composition independent hybridization in tetramethylammonium chloride, Proc. Natl. Acad. Sci. USA 82:1585–1588.

RAPID ISOLATION OF DNA FROM BODY FLUIDS USING A NOVEL NUCLEIC ACID CAPTURE REAGENT

J. Harding, G. Gebeyehu, R. Beebe, D. Simms and L. Klevan

Life Technologies Inc.
Gaithersburg, Maryland

Rapid and quantitative isolation of DNA from body fluids is an essential step in the analysis of forensic samples using DNA probes. To circumvent some of the limitations of current techniques, we have synthesized and characterized the properties of, a novel nucleic acid capture reagent consisting of a methidium moiety bound to a sepharose bead by a spermine linker. Using this reagent, DNA can be isolated rapidly from body fluids without the use of organic solvents and can be characterized by techniques such as hybridization, amplification and sequencing.

MATERIALS AND METHODS

Synthesis of the reagent is described in detail elsewhere (J. Harding *et al.*, submitted). A typical sample preparation protocol is as follows: Fifty μl of capture reagent suspended in water was added to 11 μl of proteinase K treated human serum in a microcentrifuge tube. The sample was placed on an end-over-end laboratory rotator for 30 min. at room temperature. The tube was spun for 30 sec. in a microcentrifuge, and the supernatant was removed. The pelleted capture reagent-DNA complex was resuspended in 100 ml of 0.5 N NaOH on a rotator for 10 min. and spun as before. The supernatant, containing the DNA released from the capture reagent, was analyzed by dot blotting or by the Polymerase Chain Reaction (using the Gene-Amp Kit, Perkin Elmer-Cetus, Norwalk, CT). The DNA was isolated from whole blood, human cell cultures and M13 phage lysates by minor variations of this protocol.

RESULTS

Initial experiments in which radioactive DNA was captured from buffer solutions or human serum indicated that capture of DNA by the reagent is independent of the salt concentration and pH of the sample, the presence of protein denaturants and detergents and the sample volume (μl to ml). Thus, large differences in sample composition can be tolerated. Radioactive protein added to serum was not bound by the reagent, provided that the sample was pre-treated with proteinase K.

Capture and dot blot analysis of a specific DNA sequence (Hepatitis B virus DNA) isolated from human serum is shown in Figure 1 (refer to legend). A 0.5 pg of plasmid DNA containing a cloned HBV genome was captured and detected using a radioactive RNA probe and either nylon or nitrocellulose membranes (row c, columns 2 and 4). There is no signal in row f of columns 1-4 containing samples from serum lacking HBV DNA. The results depicted in lanes 7-9 of Figure 1 demonstrate that DNA can be captured from Hepatitis B virus particles present in serum and quantitated by hybridization. Signal was obtained from undiluted infected serum (columns 8 and 9, row a) and from a 1:10 dilution of infected serum with normal serum (row b) but not from normal serum alone (row f). From these data, we estimate that 50 ml of this particular serum contain about 15 pg of HBV DNA. The data shown in Figure 1, columns 5 and 6, indicate that DNA can be captured and quantitated from urine.

The results shown in Figure 2 indicate that human genomic DNA can be captured from whole blood and analyzed using the Polymerase Chain Reaction (Scharf *et al.* 1986) and primers specific for the human beta globin gene (refer to Legend). Lane 3 shows that the expected beta globin gene amplification product obtained from DNA captured from 1 ml of whole blood can be readily visualized by ethidium bromide staining. Lane 8 shows the reaction product obtained from DNA isolated from 100 cultured human HeLa cells.

CONCLUSIONS

We have shown that a nucleic acid capture reagent consisting of an intercalating moiety bound to a solid support can be used to isolate DNA from body fluids rapidly and specifically. Because of its ability to capture DNA from a wide variety of samples and its ease of use, the reagent should be of utility to forensic scientists performing DNA analysis on body fluid samples. This capture reagent is the subject of patent applications by Life Technologies Inc.

REFERENCES

Scharf, S. J., Horn, G. T. and Erlich, H. A. (1986). Direct cloning and sequence analysis of enzymatically amplified genomic sequences, Science 233: 1076–1078.
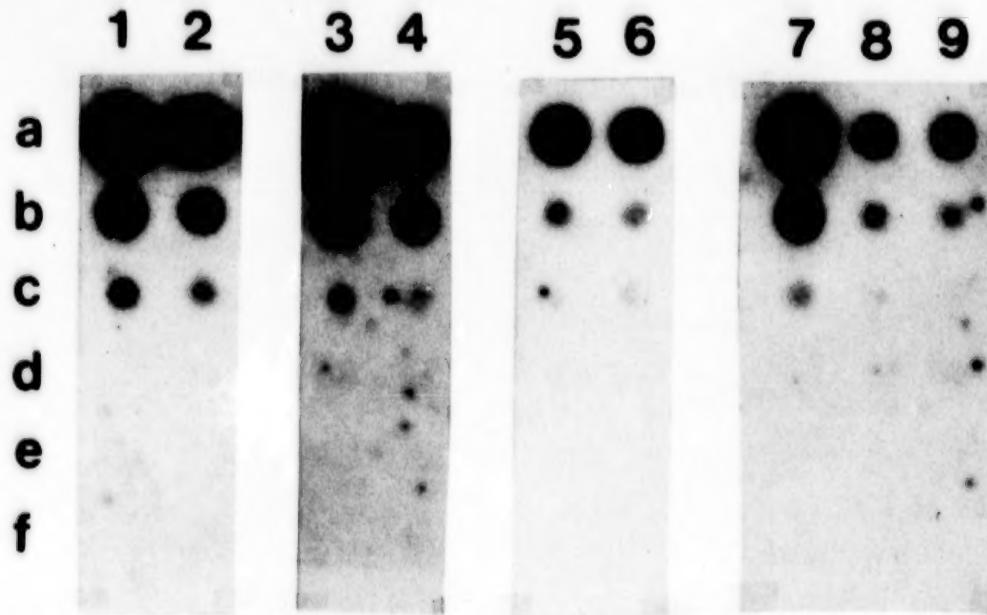
Figure 1. Dot blot analysis of viral DNAs captured from human serum and urine

Column 1. Standard amounts of a plasmid DNA containing a cloned hepatitis B virus sequence were added to 200 ml of 0.5 N NaOH and applied directly to a nylon filter. Column 2. Standard amounts of plasmid DNAs were added to normal human serum, captured and released as described in the text and applied to the filter. The filter was hybridized with a P^{32} labelled HBV RNA probe (10^7 CPM). Row a, 50 pg HBV target; b, 5 pg HBV target, c, 0.5 pg target; d, 0.25 pg target; e, 0.05 pg target, f, no target. A 7 day autoradiograph is shown.

Columns 3 and 4. The same as Columns 1 and 2, respectively, except that a nitrocellulose filter was used. A 4 day exposure of the autoradiograph is shown.

Column 5. As for Column 1 except that a plasmid DNA containing a cloned cytomegalovirus sequence was applied to a nylon membrane. Column 6. Standard amounts of CMV plasmid DNAs captured from human urine. Row a, 67 pg target; b, 6.7 pg target; c, 0.67 pg target; d, o.33 pg target; e, 0.067 pg target; f, no target. A 5 day exposure of the autoradiograph is shown.
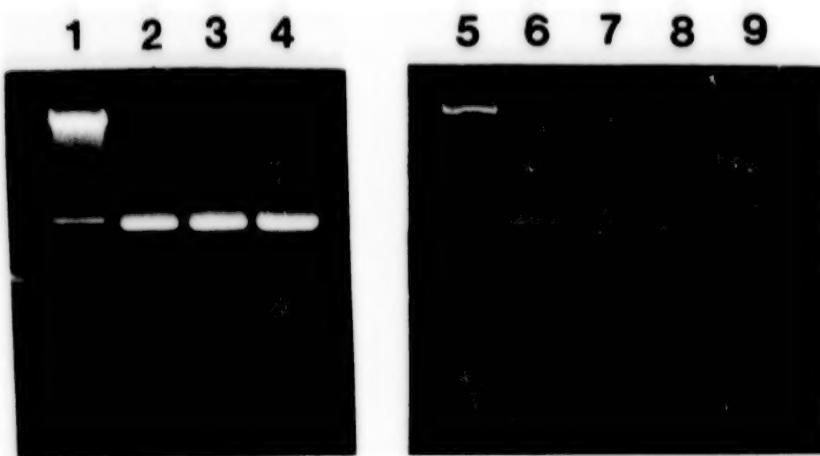
Figure 2. Agarose gel electrophoresis of products of the polymerase chain reaction performed on captured human genomic DNA.

Using the capture reagent, human genomic DNA was isolated either from whole blood or from a HeLa cell culture, the polymerase chain reaction was performed using the GH18 and GH19 primers complementary to specific human beta globin gene sequences (Scharf *et al.* 1986), and the products were subjected to electrophoresis on an agarose gel and stained with ethidium bromide.

Lane 1, the 123 bp ladder (BRL) size markers. The fragment of greatest mobility is 123 bp in size. The PCR reaction products from DNA isolated from 10 ml of whole blood (lane 2), one ml of whole blood (lane 3), from a control plasmid containing the beta globin sequence (lane 4). Lane 5, 123 bp ladder size markers. The PCR reaction products from DNA isolated from 10,000 HeLa cells (lane 6), 1,000 HeLa cells (lane 7), 100 HeLa cells (lane 8) and control plasmid (lane 9).

Lanes 1–4 and 5–9, respectively, are from different gels.

BLANK PAGE

DNA FINGERPRINTING WITH M13mp8 RF BACTERIOPHAGE USING NONRADIOACTIVE METHODS

L. Tilzer and R. Moreno

Kansas University Medical Center
Kansas City, Kansas

F. Booth

Kansas City Police Department
Kansas City, Missouri

The DNA fingerprints are Southern blots, which have a pattern resembling bar codes. The pattern is created by DNA probes that bind to variable length repeated sequences of human genomic DNA digested with restriction endonucleases. The use of wild type M13 phage DNA to detect highly variable minisatellite sequences (DNA fingerprints) was reported initially by Vassart *et al.* (1987). The ability to detect the hypervariable sequences can be attributed to a sequence present in the protein III gene of the M13 bacteriophage, in a region between coordinates 1013 and 2528 on the M13 map. Devor *et al.* (1988) modified Vassart's technique, using the 849 and 309 base pair fragments of M13 digested with *Hae*III as the gene probe to identify cultured cell lines. Improvement in the fingerprint has been claimed by Westneat *et al.* (1988) using bovine serum albumin in place of blotto during hybridization.

All the above studies have used radioactive labeled M13 as probes. The basic methodology to perform the DNA fingerprinting was with biotin labeled M13 instead of P^{32}. Using nick translated biotin labeled probes, the best DNA fingerprints were obtained under the following conditions: 41 hr. electrophoresis at 1.5 cm, alkaline transfer by Southern blot to Sure Blot (Oncor, Gaithersburg, MD) nylon, high stringency hybridization conditions and low stringency posthybridization washes.

In the fingerprints produced by this methodology there were in the 4–23 Kb size range, 15–20 bands that were unique and reproducible for each individual. Close relatives showed clear band sharing as would be expected. We have investigated the lower limits of detection of genomic DNA. We can easily visualize a fingerprint with as little as 2 μg of human genomic DNA. Since M13 is so universally available, the assay described can be applied safely and easily to criminal investigations, forensic pathology, parentage testing and many other aspects of medicine.

REFERENCES

Devor, E. J., Ivanovich, A. K., Hickock, J. M. and Todd, R. D. (1988). A rapid method for confirming cell line identity: DNA fingerprinting with a minisatellite probe from M13, Biotech. 6:200–201.

Vassart, G., Georges, M., Monsieur, R., Brocas, H., Lequarre, A. S. and Christophe, D. (1987). A sequence of M13 phage detects hypervariable minisatellites in human and animal DNA, Science 235:683–684.

Westneat, D. F., Noon, W. A., Reeve, H. R. and Aquadro, C. F. (1988). Improved hybridization conditions for DNA fingerprints probed with M13, Nucl. Acids Res. 16:4161.

BLANK PAGE

EXTRACTION OF GENOMIC DNA FROM STAINS WITHOUT ETHANOL PRECIPITATION

F. Booth

Kansas City Police Department
Kansas City, Missouri

L. Tilzer and R. Moreno

Kansas University Medical Center
Kansas City, Kansas

High molecular weight DNA is necessary to obtain DNA fingerprints and DNA types from any given sample. The method of extraction should maximize the size recovered DNA. The bulk of the recovered DNA should exceed 20 Kb. Compounding the difficulty of samples involved in forensic work are usually stains that contain small amounts of DNA. Therefore, the optimum extraction method should maximize yield and minimize shearing.

MATERIALS AND METHODS

A procedure for isolating DNA from stains using the product GENE CLEAN™ (Bio 101 Inc., La Jolla, CA) was investigated. A stain is placed in a 1.5 ml microcentrifuge tube followed by 50 ml each of a SDS/Urea lysing solution (Oncur, Gaithersburg, MD) and a Proteinase K solution. After 1 hr. at 60° C the liquid is centrifuged free from the stained material. Deionized water is added until the total volume is 150 ml. Phenol/chloroform extractions follow with the aqueous layer isolated. Saturated NaI (provided in the kit) is added at 2.5 times the recovered volume. The glass milk is then added (5 ml), with the entire solution gently agitated for about 15 min. This process permits the DNA to bind to the glass milk while other impurities remain in the solution. After the glass milk is isolated by centrifugation, the liquid can then be discarded. The pellet is rinsed twice with a cold solution (50/50 ethanol and TE buffer) for about 10 min. each. After rinsing, the pellet is vacuum desiccated to dryness. The DNA bound to the glass milk can be recovered easily into water or the necessary enzyme digestion solutions.

RESULTS

Paralleling this procedure against ethanol precipitation (Baechtel *et al.* 1988; Gill *et al.* 1987; Maniatus *et al.* 1982; McCabe *et al.* 1987) showed that Gene Clean™ can yield higher recoveries of DNA. A parallel study from bloodstains is illustrated in Figure 1. Despite the mechanical binding of DNA to the glass milk, no appreciable shearing was observed. It was learned that the step to recover DNA from the glass milk was critical in attaining high recoveries. Removal involves dispersing the pellet into the desired solution, where in a low salt medium the DNA is released from the glass milk and becomes solvated. The glass milk can then be repelleted, by centrifugation, followed by another pellet dispersion. Repeating these pelleting and dispersion steps 3 or 4 times will ensure maximum recovery and/or digestion of DNA. Finally, it was found that the glass milk should not be included with the sample for electrophoresis, because the glass milk has the effect of retarding band migration. Despite this drawback, DNA can easily be recovered in as little as 10 ml without containing any of the glass milk.

Using the product GENE CLEAN™ in an extraction procedure, we have demonstrated that high molecular weight DNA can be isolated from small samples. Restriction enzyme digestion can be done directly on the glass milk while avoiding a resuspension step. At minimal cost, isolation and digestion of DNA from a given sample can be accomplished easily within 4 hr.

REFERENCES

Baechtel, F. S. (1988). Recovery of DNA from human biological specimens, Crime Lab Digest 15:95–96.

Gill, P., Lygo, J. E., Fowler, S. J. and Werrett, D. J. (1987). An evaluation of DNA fingerprinting for forensic purposes, Electrophoresis 8:38–44.

Maniatus, T., Fritsch, E. and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

McCabe, E. R. B., Huang S., Seltzer, W. K. and Law, M. L. (1987). DNA microextraction from dried blood spots on filter paper blotters: potential applications to newborn screening. Hum. Gene. 75:213–216.
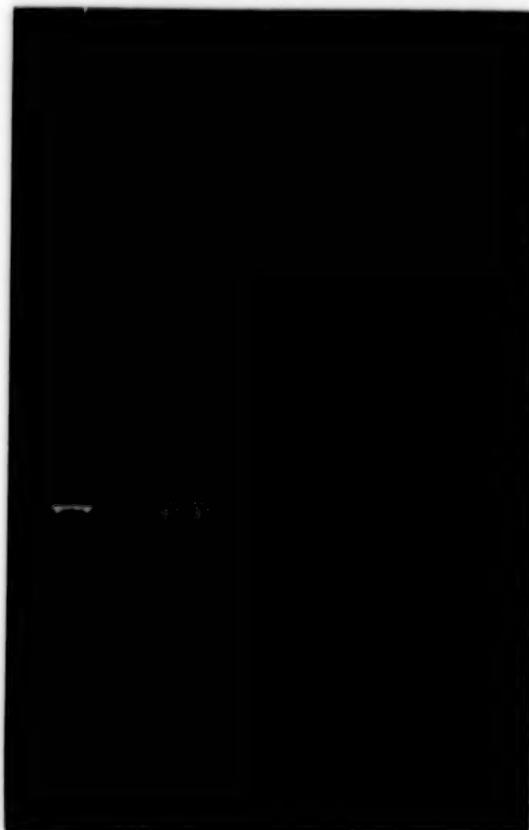
Figure No. 1. ELECTROPHORESIS OF UNDIGESTED DNA RECOVERED FROM BLOODSTAINS. This figure shows an electrophoretic run on a minigel followed by ethidium bromide staining. The furthest left land is a HindIII digest of lambda bacteriophage DNA. The lane to the right is 50 ng of undigested lambda. The next three lanes are of undigested DNA recovered via Gene Glean from bloodstains of 20, 10 and 5 ml respectively. The last three lanes are of undigested DNA recovered by ethanol precipitation from bloodstains of 20, 10 and 5 ml respectively.

CASE EXAMPLES USING DIFFERENTIAL EXTRACTION PROCEDURES

G. Herrin, Jr., R. W. Cotton, A. C. Corey, K. David, T. A. McNeil, K. R. Rubenstein, L. A. Tonelli and D. D. Garner

Cellmark Diagnostics
Germantown, Maryland

The differential extraction procedure described by Gill *et al.* (1985) is used in casework when samples contain mixtures of sperm and other tissue. Use of differential lysis can facilitate the analysis of what would otherwise be 2 overlapping DNA patterns. Although overlapping patterns can be distinct and easy to analyze, several factors can make this analysis more difficult. Some of these factors are: degradation of DNA, DNA from separate individuals present in different quantities, closely spaced bands and the presence of DNA patterns from related individuals as in cases of incest. The potential disadvantage is the possible DNA loss during the extraction procedures. Additionally, whether doing a direct or differential extraction, incomplete DNA extraction from the original sample can occur. To examine the question of whether any cellular material remains on the evidence sample following differential extraction, we have incorporated a third extraction of the evidence with the same buffer used in extract 2 (the extraction for sperm DNA).

MATERIALS AND METHODS

Differential lysis of evidentiary samples was performed according to Gill *et al.* (1985) with the following exception: buffer composition used for extracts 2 and 3 was 10 mM TRIS pH 7.6, 1 mM EDTA, 10mM NaCl, 1% SDS, 40 mM DTT and 280 μ/ml PK. Samples were incubated overnight at 56° C for all 3 extraction steps. The DNA was digested with *Hinf*I (New England Biolabs, Beverly, MA), and the resulting fragments separated by electrophoresis on a 0.7% agarose gel in TBE. Hy-bridization with the 4 single locus probes (SLPs) MS1, MS31, MS43 and G3 was executed initially with a mixture of all 4 probes followed by hybridization with each SLP individually (Wong *et al.* 1987). Autoradiographic exposure times varied depending on the amount of DNA in each sample.

The utility of the differential lysis procedure combined with a third extraction step can be illustrated by examining case examples. In many instances DNA degradation from non-sperm tissue occurs more rapidly than degradation of sperm DNA. Without prior separation of these 2 patterns, the DNA pattern from sperm would be masked by the degraded non-sperm DNA's hybridization signal.

Substantial amounts of cellular material can remain on the evidence substrate after the second extraction. Use of the third extraction retrieves this DNA thus maximizing the total amount of DNA extracted from the sample. Although combining extract 3 DNA with extract 1 or extract 2 DNA would be possible, this procedure has not proved to be helpful since either sperm or non-sperm DNA may be present in extract 3. Case examples showing the variation in amount and source of DNA remaining on evidence after extract 2 illustrate this phenomenon.

REFERENCES

Gill, P., Jeffreys, A. J. and Werrett, D. J. (1985). Forensic application of DNA fingerprints, Nature 318:577–579.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. J. (1987). Ann. Hum. Genet.

BLANK PAGE

EFFECT OF NONOXYNOL-9 ON RECOVERY OF DNA FROM SPERM

K. T. Sheridan and R. W. Cotton

Cellmark Diagnostics
Germantown, Maryland

I. M. Foster

Baltimore County Health Department
Towson, Maryland

Nonoxynol-9, a chemical ingredient of spermicides, has been shown to inhibit infection with HIV *in vitro* at concentrations as low as 0.05% (Hicks *et al.* 1985). It has been suggested that treatment with a spermicide immediately after a sexual assault would be a potential prophylactic therapy against possible HIV infection. This advantage must be weighed against the effect of nonoxynol-9 and other components of the spermicide on the ability to obtain sperm cell DNA suitable for analysis. The following experiment was designed to examine the effect of spermicides containing nonoxynol-9 on the ability to analyze DNA from sperm cells.

MATERIALS AND METHODS

One half application (about 20 ml foam volume) of foam spermicide containing 8.0% nonoxynol-9 (Emko contraceptive foam, Schering Corporation, Kenilworth, NJ) was mixed with 50 μl of fresh semen. Aliquots thus prepared were incubated at 37° C for 0, 2, 18, 24, 36, 48 and 120 hr. Untreated control aliquots of fresh semen were incubated for the same times. Following incubation, DNA was extracted from samples using a modification of Gill *et al.* (1985) and then analyzed for both the presence of high molecular weight DNA and for the ability to yield DNA band patterns by analysis using the 4 single locus probes msl, ms31, ms43 and g3 (Wong *et al.* 1987).

RESULTS

High molecular weight DNA was isolated from both control and treated semen samples, although in general the treated samples showed a poorer yield and higher degree of degradation. It is likely that the reduced yield from the treated samples resulted in part from loss of DNA during extraction (due to the detergent content). All samples, however, gave an interpretable DNA banding pattern when analyzed with the panel of single-locus probes. The results of this study show that *in vitro* treatment of semen with nonoxynol-9 containing spermicides does not prevent DNA analysis. We wish to next examine the *in vivo* effects of spermicide treatment on DNA sperm cells analysis.

REFERENCES

Hicks, D. R., Martin, L. S., Getchell, J. P., Heath, J. L., Francis, D. P., McDougal, J. S., Curran, J. W. and Voeller, B. (1985). Inactivation of HTLV-III/lav-infected cultures of normal human lymphocytes by nonoxynol-9 *in vitro*, Lancet 2:1422–1423.

Rietmeijer, C. A. M., Krebs, J. W., Feorino, P. M. and Judson, F. N. (1988). Condoms as physical and chemical barriers against human immunodeficiency virus, JAMA 259:1851–1853.

Gill, P., Jeffreys, A. J. and Werrett, D. J. (1985). Forensic application of DNA fingerprints, Nature 318:577–579.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. J. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA, Ann. Hum. Genet. 51:269–288.

BLANK PAGE

CHARACTERIZATION OF MS51. A SINGLE LOCUS VNTR PROBE WHICH PROVIDES A QUALITY CONTROL PROCEDURE TO DEMONSTRATE COMPLETE DIGESTION WITH RESTRICTION ENZYMES

A. Cawood, M. Webb, J. Riley, R. Mead, A. F. Markham and J. C. Smith

Cellmark Diagnostics
Abingdon, England

OX14 1DY and ICI Diagnostics
Cheshire, England

Current digestion controls show whether or not restriction enzyme is working on the control DNA sample but cannot show whether any given DNA sample has been digested to completion. A true digestion control should indicate digestion sufficiently enough to ensure full resolution of loci variants used for identity matching. A locus in the human genome has been discovered which is refractory to digestion with *Hinf*I. This locus's use in determining the digestion state of test loci is described below.

MATERIALS AND METHODS

Extraction of DNA from blood, electrophoresis, transfer to nylon membranes and hybridization conditions have been described previously (Jeffreys *et al.* 1985; Wong *et al.* 1987). Probes 33.15, MS1, MS31, g3 and MS43 were provided by A. Jeffreys, (Jeffreys *et al.* 1985; Wong *et al.* 1987). Probe MS51 was isolated by cloning a 3.2 Kb *Eco*R1-*Sau*3A fragment of human genomic DNA into pUC13 (Jeffreys, unpublished). Probes were labelled with a modification of Feinberg and Vogelstein's (1983) random priming method modification.

RESULTS

When compared, DNA fingerprints, produced by hybridization with multilocus probe 33.15 after digestion of human genomic DNA with *Hinf*I for periods between 20 min. and 36 hr., reveal patterns of bands which change as the digestion duration increases until a constant pattern is produced. Reprobing with the specific locus probe MS51 shows that the constant multilocus pattern is associated with 1 or 2 bands. Thus, the MS51 locus is refractory to digestion with *Hinf*I. Completion of digestion at this locus coincides with the final, unchanging multilocus DNA fingerprint production. Further reprobing with the specific locus probes MS1, MS31, MS43 and g3 (Wong *et al.* 1987) reveals that these loci are not refractory to digestion with *Hinf*I, and that MS51 digests to completion after these loci, which are used in identity testing.

The resistance of MS51 to rapid digestion with *Hinf*I has been demonstrated in 55 individuals selected randomly from the UK population and 25 individuals from the Indian sub-continent. No individuals have been found whose DNA the MS51 locus digests to completion before MS1, MS31, MS43 and g3.

Thus MS51 provides a digestion control for the specific DNA samples being investigated and has been used to define digestion conditions ($8\mu/\mu g$ DNA for not less than 16 hr.) which provide complete test loci digestion.

The MS51 locus characterization is in progress. The repeat unit and its flanking sequences have been isolated from the DNA of 1 individual and cloned. Restriction maps of the flanking sequence have been produced, and the region surrounding the *Hinf*I sites will be sequenced.

REFERENCES

Feinberg, A. P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity, Anal. Biochem. 132:6–13.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985). Hypervariable minisatellite regions in human DNA, Nature 314:67–73.

Wong, Z., Wilson, V., Patel, I., Povey, S. and Jeffreys, A. J. (1987). Characterization of a panel of highly variable minisatellites cloned from human DNA, Ann. of Human Genet. 51:268–288.

BLANK PAGE

PERSONAL IDENTIFICATION FOR JAPANESE USING VARIABLE NUMBER OF TANDEM REPEAT (VNTR) LOCI

K. Kasai, H. Mukoyama, Y. Nakamura and R. L. White

Howard Hughes Medical Institute Research Laboratories
Salt Lake City, Utah

K. Kasai and I. Sakai

Laboratory of Forensic Serology
Tokyo, Japan

In order to define the appropriate DNA markers for the individual identification in the Japanese population, we have characterized isolated DNA from nearly 50 unrelated Japanese people with 4 VNTR markers; YNH24 (D2S44), MLJ14 (D14S13), YNZ22 (D17S30) and JCZ3.1 (D19S20) (Nakamura *et al.* 1987).

Five μ each of DNA isolated from nearly 50 unrelated Japanese was digested with 25 units of *Hinf*I and separated by an electrophoresis in a 20 cm long, 1.2% agarose gel. A 123-base-pair ladder marker (Bethesda Research Laboratories, Bethesda, MD) was used as a molecular weight standard. After electrophoresis, DNA was transferred to a charged nylon membrane by a NaOH method. Transferred membranes were prehybridized overnight at 42° C in 5 x SSC, 20 mM NaHPO4 (pH 6.5), 10 x Denhardt's solution, 5% dextran sulfate, 0.2% SDS, 0.2 μ/ml human placental DNA and 50% formamide. For hybridization, we tested 4 DNA probes; YNH24 (D2S44), MLJ14 (D14S13), YNZ22 (D17S30) and JCZ3.1 (D19S20). Each probe mixed with a 123 base pair ladder marker in a ratio of 2:1 (w/w) was labeled with alpha-P[32]-dCTP by a oligo-labeling method. Hybridization was performed at the same conditions as prehybridization except 1 x Denhardt's solution was used instead of 10 x Denhardt's solution. After hybrid-ization, membranes were washed once in 2 x SSC at room temperature and twice in 0.1 x SSC and 0.1% SDS at 65° C. Then autoradiographs were exposed overnight at -70° C. The position of each band on the exposed film was measured manually, and the size was calculated by the least square method using the ladder marker as a standard.

The YNH24 has shown a wide distribution of alleles with the number of repeats from 10 to 85 and 96% heterozygosity, and MLJ14 has shown a wider distribution of alleles with a size range between 1.8 Kb and 95% heterozygosity. The distribution of alleles with JCZ3.1 was between 1.9 Kb and 4.4 Kb, and the heterozygosity was 77%. The YNZ22 had a 85% heterozygosity with 20 alleles between 770 b and 2.1 Kb.

In conclusion, these 4 DNA markers will be useful for the individual identification for the Japanese population.

REFERENCES

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E. and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping, Science 235:1616–1622.

BLANK PAGE

DNA TYPING USING A PANEL OF VNTR PROBES

J. S. Waye, G. G. Shutler, N. Monteith, L. Bishop and R. M. Fourney

Royal Canadian Mounted Police
Ottawa, Canada

The DNA typing procedure used at the R.C.M.P. Central Forensic Laboratory represents a compilation of protocols from many different disciplines. Genomic DNA is isolated from body fluid stains using Proteinase K digestion, successive extractions with organic solvents and ethanol precipitation. Since DNA extracted from forensic specimens may be contaminated with nonhuman DNA, DNA of human origin is specifically quantified using a simple and sensitive slot blot hybridization procedure. Quantification of human DNA takes just 4 hr. to complete, requires only a small fraction of the extract and can detect <100 pg human DNA (Waye *et al.* 1989). If the amount of human DNA is not limiting, its integrity (molecular weight) is subsequently assayed by conventional gel electrophoresis and ethidium bromide staining.

MATERIALS AND METHODS

The DNA typing procedure utilizes the restriction endonuclease *Hae*III and comprises successive cycles of hybridization with probes corresponding to 5 independent VNTR loci—specifically, D1S7, D2S44, D4S139, D17S79. As a final measure, the samples are sex-typed using a probe that recognizes a highly repetitive *Hae*III fragment on the Y chromosome (DYZ1). On each membrane, genomic DNA from both a male and female are included as controls for sex-typing and to provide a unique identifier for each probe. Lastly, the membranes are probed pMGB7 (Waye *et al.* 1987), a probe that recognizes a monomorphic 2731 bp *Hae*III fragment. Because the actual size of this fragment has been established by nucleotide sequencing, it serves as a measure for the overall precision and accuracy of the typing system. In addition, the monomorphic marker can be used to delineate any band shifts due to electrophoretic anomalies (that is, to establish that differences in RFLP patterns are due to band shifting and to establish which of the samples has migrated correctly).

RESULTS

Allele frequency population databases have been compiled for each of the 5 VNTR loci (see above). These databases consist of 700 Caucasians (Ottawa region) and 250 Native Indians (Western Canada). Efforts are ongoing to expand the Caucasian database and develop additional databases to represent various regional, racial and ethnic subpopulations. Because band shifts are not apparent when DNA profiles of unrelated individuals are analyzed, the accuracy of the database information has been independently assessed using the monomorphic D7Z2 marker. In an effort to be conservative and compatible with other North American forensic laboratories, the databases have been compiled using 31 fixed size bins based on a fragment size ladder (Lifecodes).

REFERENCES

Waye, J. S., England, S. B. and Willard, H. F. (1987). Genomic organization of alpha satellite DNA on human chromosome 7: evidence for two distinct alphoid domains on a single chromosome, Mol. Cell. Biol. 7:349–356.

Waye, J. S., Presley, L., Budowle, B., Shutler, G. G. and Fourney, R. M. (1989). A simple and sensitive method for quantifying human genomic DNA in forensic specimen extracts, Biotechniques (in press).

BLANK PAGE

IDENTIFICATION OF INDIVIDUALS WITH MULTI AND SINGLE LOCUS PROBES: THE EFFECT OF GENETIC ISOLATION, AUTOLYSIS AND SOMATIC MUTATIONS

A. Sajantila, P. Helminen, C. Ehnholm and L. Peltonen

National Public Health Institute
Helsinki, Finland

P. Saukko

University of Oulu
Helsinki, Finland

S. Pakkala

University of Helsinki
Helsinki, Finland

The Finnish population represents an isolated gene pool. This type of homogeneous population could theoretically be more prone to the problems in identification with the DNA fingerprinting due to the higher frequency of shared bands among individuals. We studied DNA-fingerprints of 52 Finnish individuals using Jeffreys' multilocus probes and established the mean probability for band sharing in unrelated individuals to be 3.98 x 10^{-18}, the number comparable to those reported from more heterogenous populations, Helminen *et al.* (1988); Jeffreys *et al.* (1985). The allelic frequencies of single locus VNTR probes are currently being analyzed in the study of 200 unrelated Finnish individuals.

Differences in DNA concentration (ng/mg) in different tissues with post mortem (P.m.) periods of 38h and 151h.
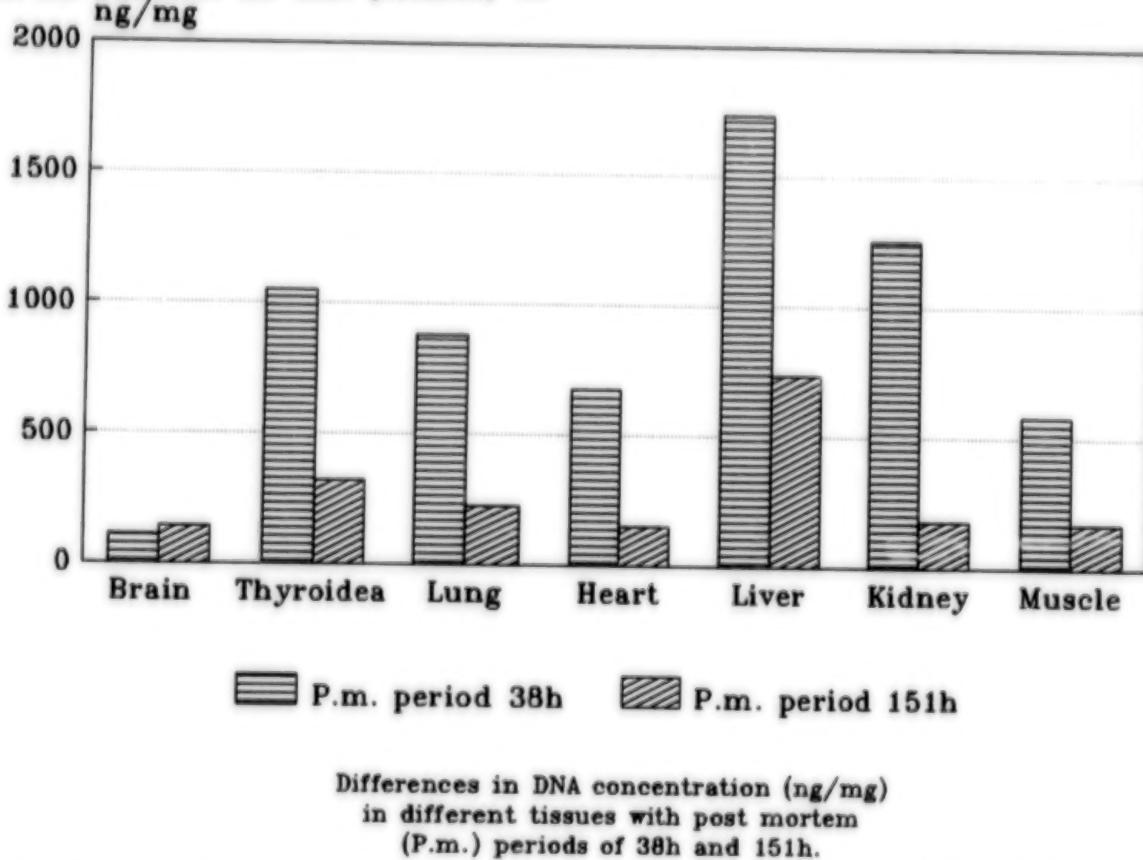
Figure 1. Differences in DNA concentration (ng/mg) in different tissues with post mortem (PM) periods of 38h and 151h.

MATERIALS AND METHODS

We studied the effect of autolysis on the amount and quality of extractable DNA from the various tissues of autopsy samples. The amount of DNA was determined by both fluorometric and spectrophotometric assay (OD 260 nm). This finding was constant from the shortest (38 hr.) to the longest (151 hr.) post mortem period of our study (Figure 1). The quality of extractable DNA was analyzed from 2–3 g DNA samples which were run on Agarose minigels in the presence of ethidium bromide. The quantity of high molecular weight DNA (>22 Kb) was highest in the samples from kidney, lungs and thyroidea. The degradation to low molecular weight DNA did not seem to be significant during the period of this study (151 hr., Figure 2). These findings are in agreement with other published data, (Bär *et al.* 1988). Also, the DNA fingerprints produced with Jeffreys' multilocus probes could be clearly identified in all tissues, even in the tissue samples from autopsies 151 hr. post mortem (Figure 3). However, a considerable loss of high molecular weight fragments in the hybridization pattern was obvious when post mortem time increased, reducing the informativeness of fingerprinting analyses of old post mortem samples. The application of a single locus VNTR probes for post mortem identification is currently being linked to the analyses of allelic frequencies of these probes (Figure 4).
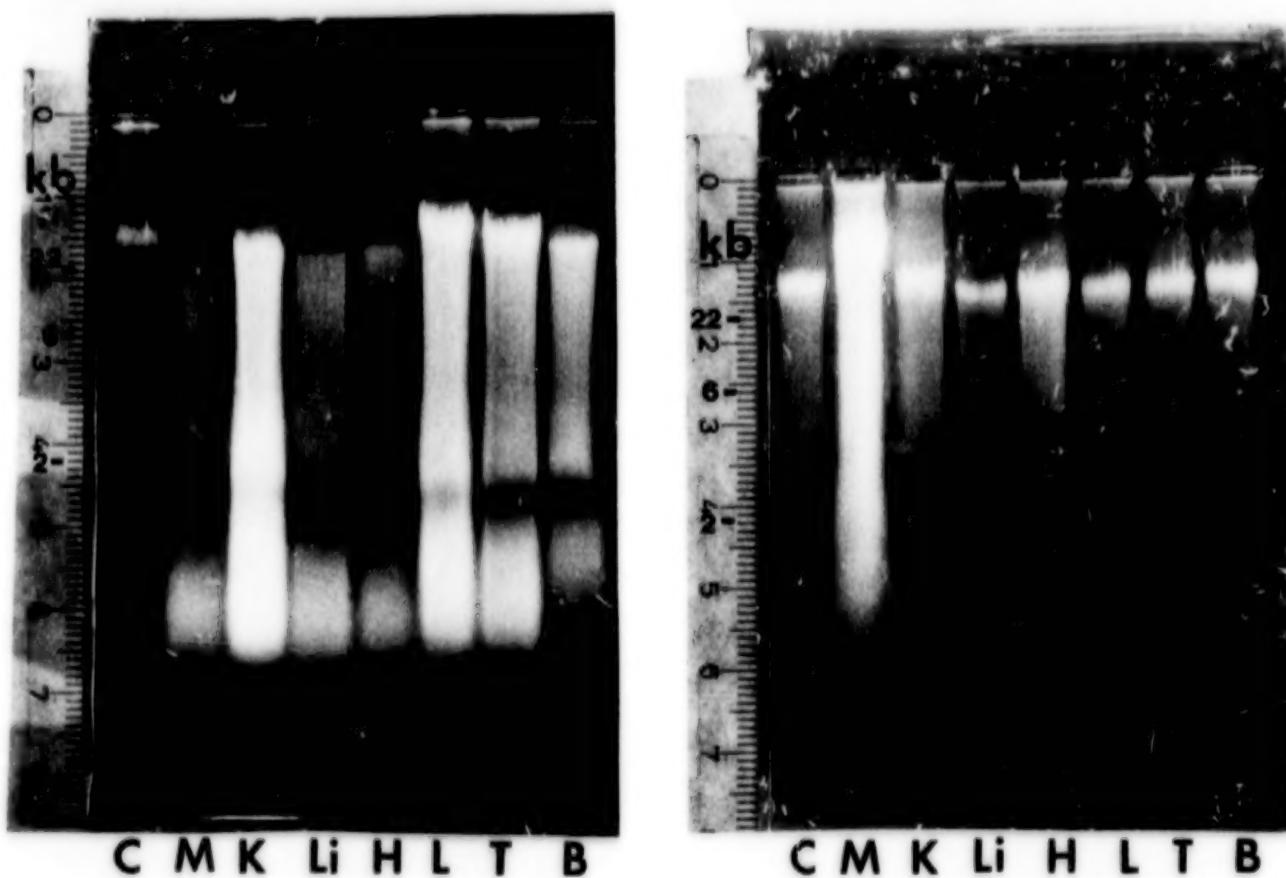
Figure 2. DNA samples (2–3 μg) run on Agarose minigels. Post mortem periods 38 hrs (left) and 151 hrs (right). (B=brain, T=thyroidea, L=lung, H=heart, Li=liver, K=kidney, M=muscle, C=control DNA)
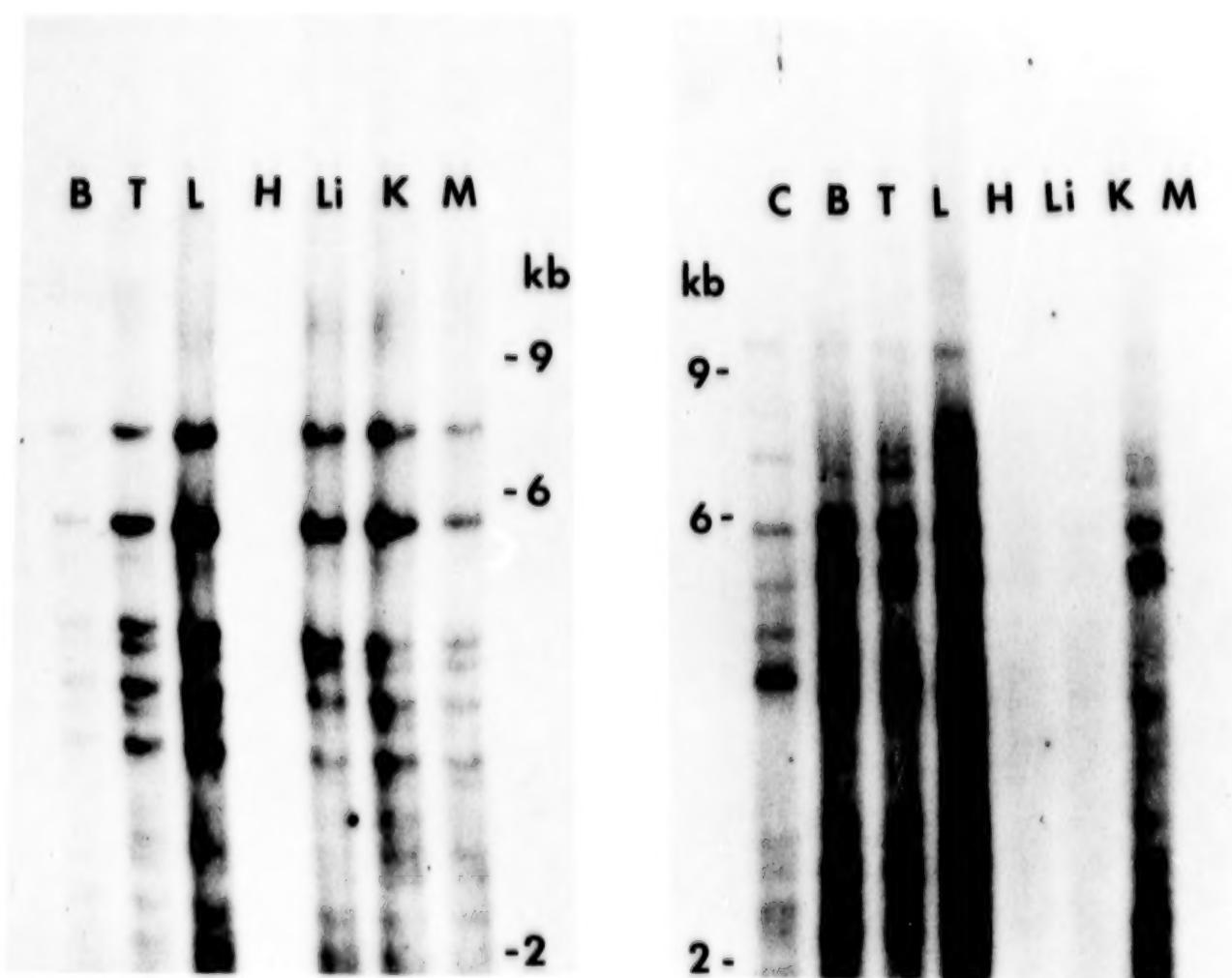
Figure 3. DNA-fingerprints produced with Jeffreys' multilocus probe (33.15). Tissue samples from 38 hrs (left) and 151 hrs (right) autopsies. (B=brain, T=thyroidea, L=lung, H=heart, Li=liver, K=Kidney, M=muscle, C=control DNA)

RESULTS

Surprisingly, the individual DNA fingerprint pattern obtained within multilocus probes is not always stable but can undergo temporal changes. In rare occasions, these changes may even have forensic relevance. We discovered distinct differences in DNA patterns obtained with multilocus probes between remission and relapse bone marrow samples of the same individual in 50% of 28 studied patients with acute lymphoblastic or myeloid leukemia. In most of these cases the differences were

also detected when wild M13 phage was used as a probe but could not be detected with single locus VNTR probes, (Pakkala *et al.* 1988). Thus the possibility of somatic mutations—although representing a rare event—should be taken into account when identification of individuals is performed using cellular material from different tissues.(33.15). Tissue samples from 38 hrs (left) and 151 hrs (right) autopsies. (B=brain, T=thyroidea, L=lung, H=heart, Li=liver, K=Kidney, M=muscle, C=control DNA)
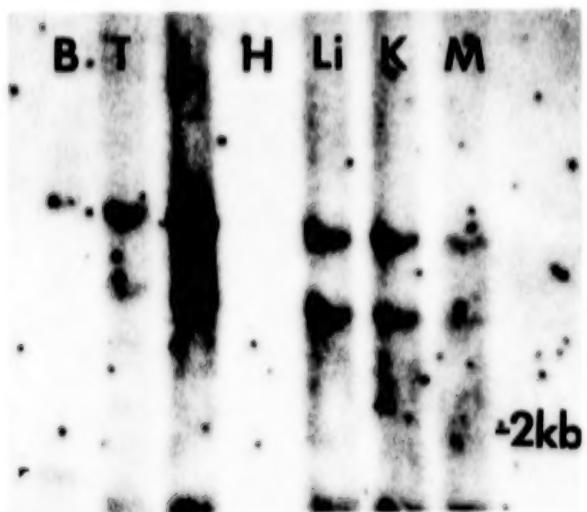
Figure 4. Hybridization with single locus VNTk probe YNH24. Tissue samples from 151 hrs autopsy. (B=brain, T=thyroidea, L=lung, H=heart, Li=liver, K=Kidney, M=muscle)

REFERENCES

Bär, W., Kratzer, A., Machter, M. and Schmid, W. (1988). Postmortem stability of DNA, Foren. Sci. Internat. 39:59–70.

Helminem, P., Ehnholm, C., Lokki, M. L., Jeffreys, A. J. and Peltonen, L. (1988). Application of DNA fingerprints to paternity determinations, The Lancet March 12:574–576.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1989). Individual specific fingerprints of human DNA, Nature 316:76–78.

Pakkala, S., Helminem, P., Saarinen, U. M., Alitalo, R. and Peltonen, L. (1988). Differences in DNA-fingerprints between remission and relapse in childhood acute lymphoblastic leukemia, Leukemia Research 12:757–762.

TRACKTEL: AN ELECTROPHORETIC PATTERN IMAGE PROCESSING SYSTEM FOR THE FORENSIC LABORATORY

H. W. J. Harding, A. M. Ross and J. C. S. Fowler

Forensic Science Centre
Adelaide, South Australia

I. Miller

Forensic Science Technology International Pty. Ltd.
The Levels, South Australia

A computerized recording system based on IBM PC/AT hardware has been developed to capture and process electrophoretic patterns. In particular, it can read, store and sort the complex multi-locus DNA patterns as well as single-locus DNA probe patterns, PCR dot blots and polymorphic protein patterns such as Hp and PGM.

The video image is captured and stored in a digital form. Contrast enhancement using pseudo-coloring ensures that weak bands can be viewed clearly. As well as storing a photographic image of the original pattern (either autoradiograph or biotin labelled), the image is converted to a block diagram and a graphic format which records position, width and intensity of the band. These converted images can be checked visually by the operator against the original patterns and quality assurance performed. This procedure ensures the accuracy of the information to be saved without altering the original information.

In the DNA mode the position of each band is automatically compared to that of standard size markers (after normalization) and expressed as size (Kb). When unknown case sample patterns are compared to the data base, the computer, with the aid of ORACLEtm Relational Database Management System, produces a list of probable matches in the order of those most likely. These images are reviewed by the operator and compared visually with the unknown in order to make a final identification.

The other key benefit gained by using ORACLE is the security of the data. ORACLE gives the system manager complete control over which individuals have:

—update privileges
—Quality Assurance privileges
—read only access

Therefore, the integrity of sensitive forensic or paternity information can be guaranteed by producing a complete audit trail logged by user I.D., time and date of any access to the system as well as identifying any records that have been altered.

The system is used to capture the image of the ethidium bromide stained gels both before and after restriction procedures. As a further quality assurance procedure the image associated case data are stored on an optical disc (write once, read many times). Hard copies for inclusion in the case files are produced via a video printer.

BLANK PAGE

EFFECTS OF VISCOSITY, PORE SIZE, IONIC STRENGTH AND POWER APPLICATION ON THE RELATIVE MOBILITY OF DNA FRAGMENTS SEPARATED IN POLYACRYLAMIDE GELS WITH DISCONTINUOUS BUFFERS

R. C. Allen and G. M. Graves

Medical University of South Carolina
Charleston, South Carolina

B. Budowle

FBI Laboratory
Quantico, Virginia

The purpose of this work was to alter the relative mobility and thus, resolution of DNA and DNA restriction fragments in polyacrylamide gel by changing physical and electrochemical parameters in the gel system.

MATERIALS AND METHODS

Separation of Polymerase Chain Reaction (PCR) amplified products, restriction fragments and DNA ladders were carried out on rehydrated 350μ thick polyacrylamide gels. Horizontal, multizonal electrophoresis (MZE) using sulfate-borate and chloride-borate discontinuous buffer systems with constant power or pulsed constant power, (PCP) Dilworth (1971, 1973) was employed on 5–10% T gels crosslinked with Bis at 3.5% C. Glycerol adduct in the rehydration medium was used to alter the gel viscosity. The borate trailing and Tris counter ions, in quantitative excess, were soaked into polyacrylamide or crosslinked agarose gel plugs. These ions were then placed onto the anodal and cathodal ends of the gel, and the electrical bridge formed by electrodes placed directly onto the surface of the gel plugs, Allen *et al.* (1989). Samples (0.5, −1.0 ml) at 1/10th ionic strength of the leading ion were placed directly onto the surface of the rehydrated gels. After the DNA migrated into the gel under low power conditions, the leading/trailing ion boundary passed through the partially separated DNA bands and zone sharpened each, Allen *et al.* (1969). Viscosity, ionic strength, gel pore size and mode of power delivery were employed individually and in combination to alter DNA mobility relative to the boundary (Rf). Separated DNA was visualized in the gel with a modified silver stain of Goldman *et al.* (1982).

RESULTS

As shown in Figures 1a and b., varying the viscosity of the gel with glycerol alters Rf and resolution potential at a given gel pore size, ionic strength and separation distance. The PCR amplified 242 base pair (bp) DQα is resolved from the 246 bp standard in conditions of increased viscosity 1b but does not even unstack in the non-viscous gel at the same ionic strength and pore size as shown in Figure 1a. Viscosity linearly effected Rf as shown graphically in bp in Figure 2.

The 5% T–3.5% C gel 20 cm in length containing 10% glycerol, shown in Figure 3b, has resolution of $+-1.6$–3.3 bp from 100 to 500 bp and $+-$ 4–7 bp from 492 and 1107 bp. Reducing viscosity at the same ionic strength and pore size, with either constant power or PCP as shown in Figures 3a and 1a, produced similar resolution in the 492 to 1107 bp region on a shorter 9 cm gel. This similarity resulted partly because of the narrower zone widths from the delayed unstacking. The DNA less than 344 bp in size did not unstack from the boundary. In Figure 4, 10,000 bp were resolved by decreasing the gel pore size and increasing the viscosity in a 10% T, 3.5% C gel with 10 glycerol. Only 5–6000 bp were resolved in a 5–6% T 3.5% % C gels. Larger DNA in 5–6 % T gels piled up and migrated well into the gel, apparently due to its high charge to mass ratio. Increasing the ionic strength also increased Rf by delaying unstacking from the boundary. As shown in Figures 1a and 3a, PCP also increased Rf in the higher bp regions. The increase was considerable over that obtained with similar levels of constant power. Separation using PCP also produced narrower DNA zones and allowed more rapid separation.
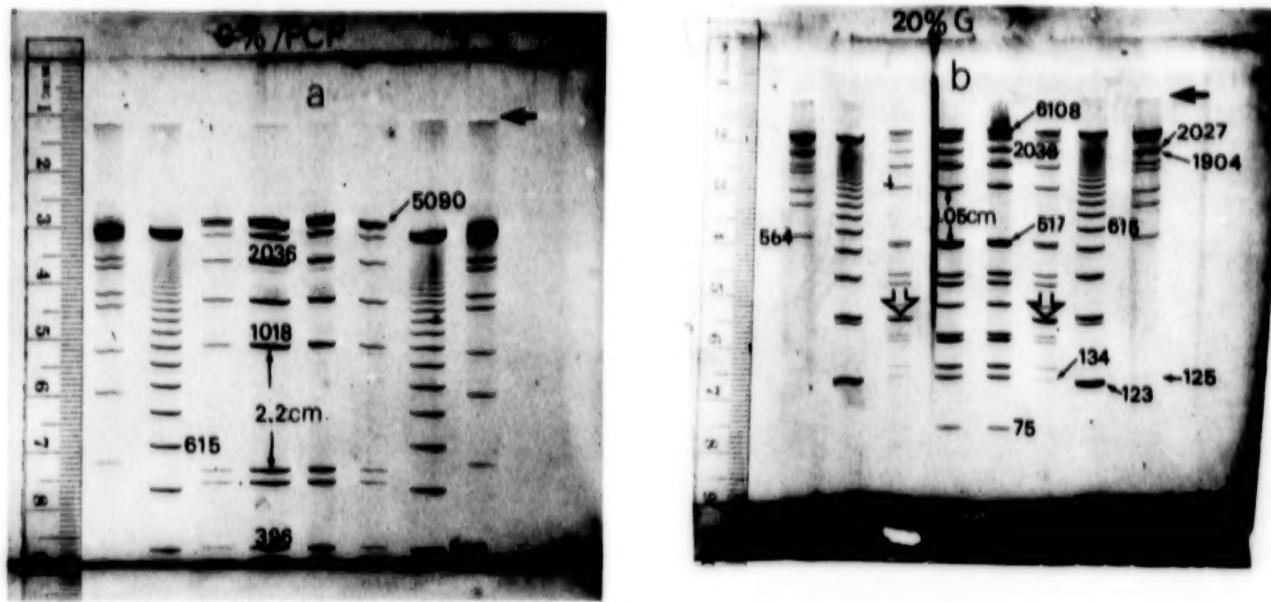
Figure 1. PCP separation in a chloride-borate system in 5% T, 3.5% C rehydratable polyacrylamide gels. Panel a: Lane 1, 0.25 *Hin*dIII/*EcoR* 1 restriction enzyme cut of lambda phage DNA; Lane 2, 62.5 ng BRL 123 bp ladder; Lane 3, 6.25 ng BRL 1000 bp ladder with DQ adduct; Lane 4, 62.5 ng BRL 1000 bp ladder; Lanes 5–8 are mirror images of Lanes 1–4. Panel b samples are the same as samples in panel a, with gel rehydrated in the presence of 20% glycerol. Large open arrows mark DQ and the solid arrows atop the sample application site. Separation times were 1 hr. and 50 and 1 hr. 55 min., respectively, using Pulsed Constant Power at 20° C.

CONCLUSIONS

Altering unstacking limits with the procedures above allow separation and resolution tailoring. Regions between $<72-10,000$ bp may be selectively closed or opened simply by altering one or more of the 5 parameters. These sieving and viscosity data indicate that electroendosmosis probably has great influence role in agarose DNA electrophoretic separation of fragments greater than 5,000 bp in size. These parameters may be useful to tailor gels for higher resolution in DNA sequencing, also.

Viscosity Effect on Rf
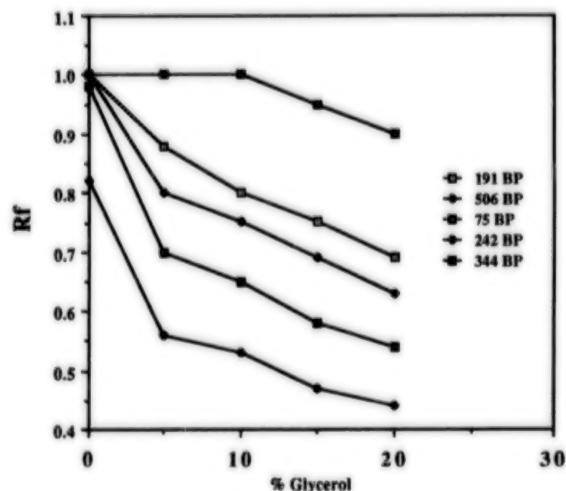
Figure 2. Mobility relative to the boundary as a function of viscosity effect with increasing amounts of glycerol in the rehydration medium.

250
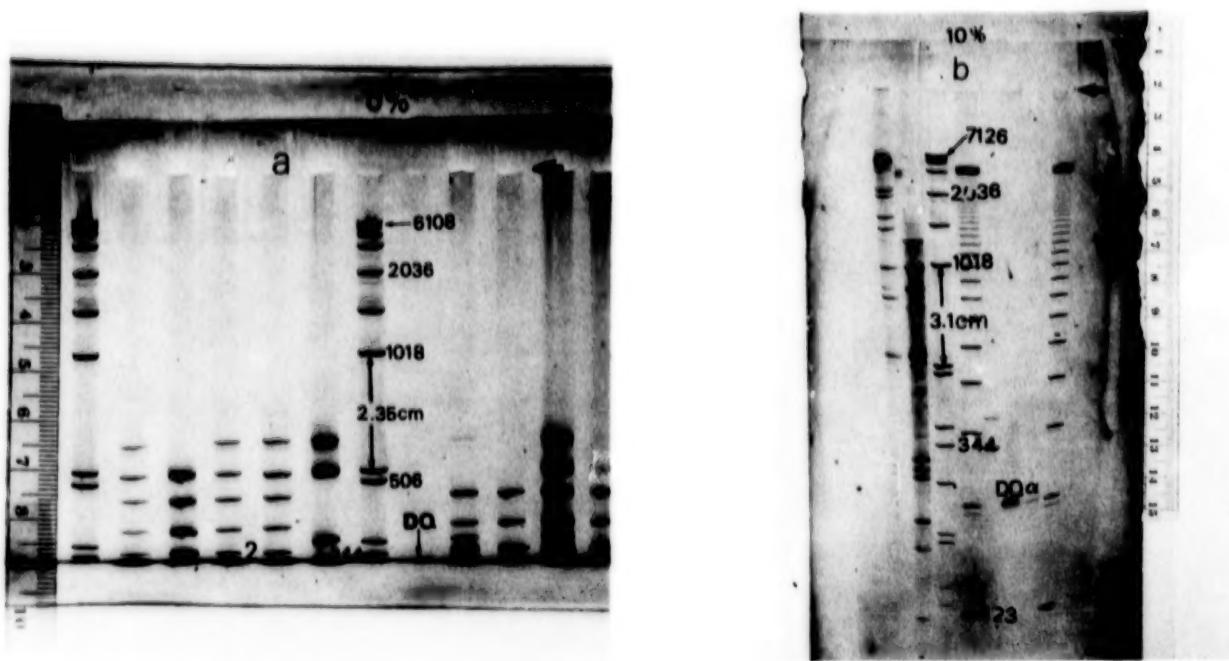
Figure 3. Panel a: A 350 ml 750 ml thick 20 cm long wedge gel 5% T 3.5 percent C polyacrylamide gel backed on GelBond Pag and hydrated in 0.038M Tris-sulfate in 10% glycerol at pH 9.0. Lane 1, 0.25 g HindIII/EcoR 1 restriction enzyme cut of lambda phage DNA; Lane 2, 0.125 g HAE restriction enzyme cut of phi x 174 phage DNA; Lane 3, 0.125 g BRL 1000 bp ladder; Lane 4. 0.125 g BRL 123 bp ladder; Lane 5, 1:10 Dilution of DQ; Lane 6, 1:100 dilution of DQ; Lane 7 repeat of lane 3. Separation times were 4 hr. and 2 hr., respectively, with ramped constant power from 1–3 Watts at 20° C. Panel b: A 5% T, 3.5% C gel hydrated in 0.038M Tris-sulphate. Lane 1, 0.125 g BRL 1000 base ladder; Lanes 2–6 PCR amplified products from Duchenne Muscular Dystrophy; Lane 7, Repeat of Lane 1; Lane 8, PCR amplified DQ not unstacked; Lanes 9–12 additional Duchenne amplified products.

REFERENCES

Allen, R. C., Moore, D. H. and Dilworth, R. H. III (1969). Polyacrylamide gel electrophoresis with discontinuous buffer systems at a constant pH, J. Histochem. Cytochem. 17:189.

Allen, R. C., Graves, G. and Budowle, B. (1989). Polymerase chain reaction amplification products separated on rehydratable polyacrylamide gels and stained with silver, BioTechniques in press.

Dilworth, R. H. III (1971, 1973). Apparatus for particle separation, U.S. Patents Nos. 3,630,882 and 3,712,859.

Goldman, D. and Merril, C. R. (1982). Silver staining of DNA in polyacrylamide gels: linearity and effect of fragment size, Electrophoresis, 3:24–26.
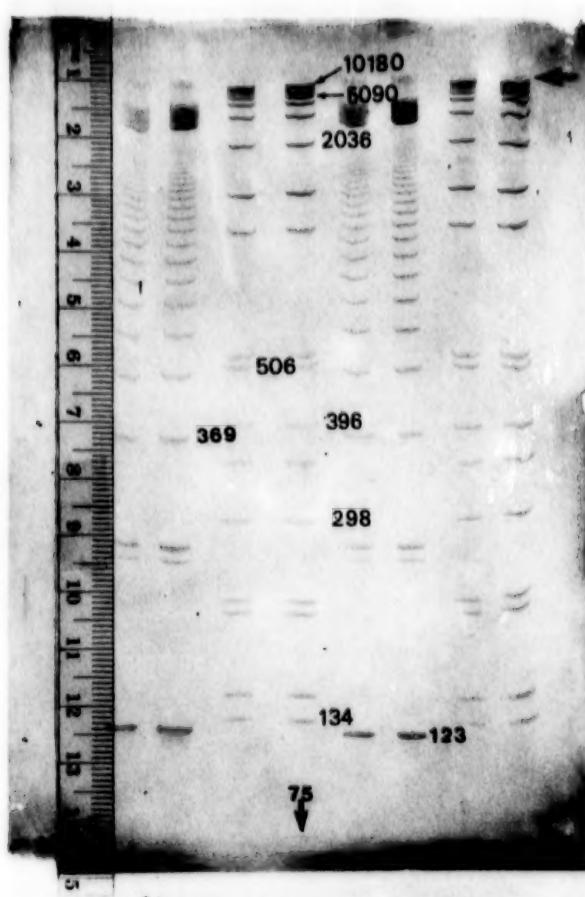
Figure 4. A 10% T, 3.5% C gel rehydrated in 0.038M Tris-sulphate at pH 9.0 with a separation distance of 14.5 cm. Lanes 3, 4, 7, 8, 0.125 g BRL 1000 bp ladder; Lane 1, 2, 5, 6 and 0.125 g BRL 123 bp ladder; showing resolution of 10,000 bp Lanes 3 and 4. Separation time 3 hr. using ramped constant power from 1 to 5 Watts at 20° C.

DESCRIPTION AND EVALUATION OF QUANTITATIVE METHODS USED TO ASSESS THE STRENGTH OF CORRESPONDENCE BETWEEN MANY-BANDED PATTERNS RESULTING FROM DNA TYPING

D. A. Stoney

University of Illinois at Chicago
Chicago, Illinois

Application of multilocus DNA probes to human DNA results in highly-individual banding patterns. Optical scanning of banding patterns from multilocus DNA probes is a convenient method of documentation. After patterns are digitized, band positions can be standardized utilizing control samples, and the data can be converted to a sequence of numbers representing adjusted band positions. Banding patterns from many case samples and many suspects may then be stored, and newly acquired samples can be compared with the samples on file. For this process a comparison algorithm is required that can be used to measure the strength of correspondence between 2 samples.

An algorithm currently used for this purpose determines the proportion of bands in each sample that correspond to the other and takes their product. Thus if one sample has N1 bands, the other sample has N2 bands and C bands correspond, the measure of correspondence is given by Equation 1.

$$(C/N1) \times (C/N2) \quad \text{Equation 1.}$$

In this paper several additional comparison algorithms are considered, and the effectiveness of each is evaluated.

Four assumptions are made:

1. All band positions are equally likely to be occupied.

2. Positions of band occurrences within the pattern are independent of one another.

3. The position about which a band is centered (or most dense) fully describe its character for comparison purposes.

4. A discrete number of possible band sites can be defined.

Under these assumptions the data resulting from a comparison of 2 digitized banding patterns if fully given by 4 variables, defined as follows:

N1, the number of bands occurring in sample 1
N2, the number of bands occurring in sample 2
C, the number of corresponding bands
S, the number of band sites

Measures of correspondence are considered that reflect the percentage of matching bands, the percentage of matching positions, and a probabilistic method estimating the chance of randomly observing a given number of matching bands. The methods were evaluated by considering the effects of variations in sample quality, analytical precision and band density.

The measure on percentages of corresponding bands is insensitive to the resolution of the system and is rejected. Measures based on percentages of band sites are also faulty, as they are insensitive to band number or equivalently, band density. The probabilistic method, based on a hypergeometric probability model, is most suitable by these criteria. However, it fails to discriminate well among patterns with substantial band correspondence. Further discrimination is achieved by considering the number of residual band discrepancies—those that cannot be explained through differences in either sample quality or assay sensitivity. For a given number of residual discrepancies, the hypergeometric model is suitable and is proposed for use.

BLANK PAGE

EFFECTS OF DNA DAMAGE AND DEGRADATION OF RFLP ANALYSIS

K. C. Konzak, R. Reynolds, C. von Beroldingen, M. Buoncristiani and G. F. Sensabaugh

University of California
Berkeley, California

Damage and degradation of DNA in forensic samples can occur by 3 major mechanisms: (1) strand breakage, (2) damage to the bases and (3) by cross-linking of bases to each other or to protein. To address the effects of DNA damage, if any, on restriction fragment length polymorphisms (RFLP) typing systems, a systematic study of DNA damage and degradation was undertaken. Naked genomic DNA was used to establish calibrated conditions for each of these damage sources. We are exploring strand breakage by physical shearing and by enzymatic cleavage, base damage by depurination with acid and cross-linking by UV using this approach.

MATERIALS AND METHODS

Strand breakage was examined both as random cleavage, with DNAase I, and as specific cleavage, with restriction enzymes, for example, *EcoRI, BamHi* and *BstNI.* Isolated DNA was subjected to DNAaseI (0.05-50pg/μg DNA, in the presence of both Mg++ and Ca++) at 37° C for timed intervals from 30 sec. to 30 min. The reactions were stopped with 0.2M EDTA, EDTA and heat, or 0.2M EDTA and 0.1% SDS together, the latter being most effective. Restriction endonucleases were used at varying concentrations (for example, 0.1-10 u *EcoRI* per up DNA) for a 2 hr. incubation at 37° C and again halted with 0.2M EDTA. These treatments produced digestion ranging from partial to complete on agarose gels with ethidium bromide staining. The degrading enzymes were removed from the samples by chloroform:phenol extraction and the partially digested or damaged DNA isolated by ethanol precipitation or by microdialysis using centricon tubes (Amicon, Danvers, MA). The damaged DNA was then digested with *HaeIII,* with the restriction fragments separated on 1% agarose by electrophoresis and blotted to Zeta-probe (Bio Rad, Richmond, CA) charged nylon membranes. The RFLP were detected by hybridization with random primed (Feinberg and Vogel-stein, 1983) single locus VNTR probes commonly in use, for example, D187, (MS-1, courtesy Cellmark, Germantown, MD), D2S44 (pYNH24, originally cloned from Ray White's lab), and D17S85 (3'HVR, courtesy Collaborative Research, Bedford, MA).

RESULTS

The VNTR probes were hybridized with unrestricted DNAase treated DNA. Hybridization was detected over a range of fragment sizes, down to a few 100 base pairs or less in length. Similar results with probes against highly repetitive elements suggest that hybridization for quantitation may not provide a good indicator of detectable VNTR fragments (that is, that size exceeds 2 Kb). Southern blots of these degraded DNA samples, after *HaeIII* restriction, show a loss of RFLP pattern as the average size of the degraded DNA decreases. With increasing degradation, the higher molecular weight band lost intensity relative to the shorter fragment. No band shifts or pattern alterations were observed. Surprisingly little background hybridization was observed, as well.

Degradation of DNA with restriction enzymes prior to *Hae*III digestion ultimately produces patterns typical of the double digest. Where other restriction sites are present within *Hae*III fragments, double digestion leads to predictably smaller fragment patterns. The presence of the restriction site is dependent on the locus and the specific fragments detected. Random changes were not observed by any enzymatic degradation of isolated genomic DNA.

REFERENCES

Feinstein, A. P. and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity, Anal. Biochem. 137:266–267.

BLANK PAGE

EFFECTS OF DNA DEGRADATION ON AMPLIFICATION BY THE POLYMERASE CHAIN REACTION

R. Reynolds, C. von Beroldingen and G. F. Sensabaugh

University of California
Berkeley, California

The DNA isolated from evidence samples is often found in a degraded state. Degraded DNA is susceptible to amplification by the polymerase chain reaction (PCR) provided some full length target sequence remains. We describe here an investigation of the consequences of DNA strand breakage on the amplification of target sequences of different sizes.

MATERIALS AND METHODS

Strand breakage has been generated in genomic DNA by enzymatic digestion using DNAase I, *Eco*RI and *Bst*NI. The DNAase I digestion is relatively independent of DNA sequence; accordingly, DNAase I cut sites will be randomly distributed. The restriction enzymes *Eco*RI and *Bst*NI, in contrast, cut DNA at specific sites. Samples exhibiting progressive degrees of degradation were produced by varying the enzymatic digestion time; the extent of degradation was assessed by agarose gel electrophoresis. Each sample was tested for amplification at 3 concentrations of starting DNA: 20 ng, 50 ng and 200 ng. The 3 target sequences for amplification were 82 bp, 242 bp and 832 bp regions of the DQα gene (Saiki *et al.* 1986); all contained the DQα type specific sequence. The samples were amplified through 30 cycles 'in 50 μl reaction volumes. Amplification of the 82 and 242 bp sequences employed 2.5 U Taq polymerase under the following cycling conditions: 30 sec. denaturation (94° C), 30 sec. reannealing (55° C), and 1 min. extension (72° C). Amplification of the 832 bp segment employed the same conditions except that 5 U Taq was used and the extension interval was 1.5 min.

RESULTS

Random strand breakage by DNAase I digestion had no effect on amplification of the 82 bp region, even when highly degraded samples were used. Partially degraded samples amplified more efficiently than undegraded samples and is probably due to more efficient strand dissociation in the partially degraded samples. As expected, amplification of the 242 and 832 bp regions diminished as the mean size of the degraded fragments approached the size of the target sequence.

Digestion with *Bst*NI, which cuts within the target sequence regions, reduced the yield of PCR products following the pattern seen with the DNAase digested samples. In contrast, digestion with EcoRI, which does not cut in any of the target sequences, had no effect on amplification. These results show that strand breakage in a template sequence blocks amplification.

To confirm that the amplification products obtained from the degraded samples were not modified with regard to type, DQα typing was done using the immobilized allele specific probe format developed by Cetus. All samples tested typed correctly.

These results demonstrate that PCR can be used for the typing of degraded DNA samples. The only constraint is that the degraded sample contains some fragments of intact target sequence. For badly degraded samples, it will be necessary to employ primer sets defining short target sequences.

REFERENCES

Saiki, R. K., Bugawan, T. L., Horn, G. T., Mullis, K. B. and `Erlich, H. A. (1986). Analysis of enzymatically amplified β-globin and HLA–DQα DNA with allele-specific oligonucleotide probes, Nature 324:163–166.

BLANK PAGE

EFFECTS OF UV DAMAGE ON DNA AMPLIFICATION BY THE POLYMERASE CHAIN REACTION

M. Buoncristiani, C. von Beroldingen and G. Sensabaugh

University of California
Berkeley, California

Exposure of DNA to ultraviolet radiation can result in several forms of DNA damage, prominently including the generation of pyrimidine dimers, the formation of DNA strand cross-links and DNA-protein cross-links. Many studies have demonstrated that the extent of DNA damage increases as the level of UV exposure increases. Evidence samples of DNA may suffer some degree of UV damage, thus it is important to know what effect UV damage might have on DNA analysis. We have investigated the effects of UV damage, specifically pyrimidine dimer formation, on the polymerase chain reaction (PCR).

MATERIALS AND METHODS

Both naked DNA and DNA in whole cells were exposed to measured doses of short (254 nm) and long (365 nm) wave UV irradiation. Purified genomic DNA was exposed in solution, and whole blood and semen were exposed in both the liquid and dried stain forms. Exposures ranged from 0 to 30,000 J/M^2. The DNA from the UV exposed samples was then amplified with standard 30 cycle PCR, using a 110 bp sequence from the β-globin gene or the 242 bp typing sequence from the HLA DQα gene as target sequences. The quantity of PCR product produced for each sample was then assessed using agarose gel electrophoresis and ethidium bromide staining.

RESULTS

It was found that, for naked DNA in solution, short wave exposures exceeding 200J/M^2 resulted in reduced PCR product; above 2000J/M^2, PCR product was lost. In contrast, there was little loss of PCR product from extensively irradiated liquid whole blood, liquid semen or blood stain material. These results suggest that the DNA in these materials is shielded from UV exposure. Irradiation of dry semen stains with short wave UV, however, resulted in greatly reduced DNA recoveries with corresponding PCR product reductions. Irradiation of DNA in solution or of body fluids with long wave UV had no detectable effect on PCR, regardless of dosage.

Analysis of primer extension products gives evidence that Taq polymerase is blocked at pyrimidine dimer sites. Samples of purified β-globin PCR fragment DNA in solution were exposed to measured doses of short wave UV. The irradiated DNA was used as the template for a 1 cycle extension using a labeled primer, and the reaction products were analyzed on a 7% polyacrylamide sequencing gel alongside standard di-deoxy sequence control lanes. The reaction products were fragments terminating at adjacent pyrimidines on the template strand. Doses up to 120J/M^2 resulted in fragments which were fairly evenly distributed along the length of the molecule. Higher doses resulted in a shift of the band profile toward shorter fragments; 7200J/M^2 resulted in only the first 7 shortest termination fragments. Above this dose there was no change.

While it is unlikely that DNA in evidence materials will be exposed to high doses of short wave UV, it is important to define parameters effecting PCR efficiency. The pyrimidine dimer can be used as a model DNA lesion. These results suggest that dimer damage can present an insurmountable obstacle to Taq polymerase, thus blocking DNA amplification by PCR.

BLANK PAGE

A COMPARATIVE STUDY OF DNA EXTRACTED FROM SEVEN POSTMORTEM TISSUES

S. Swarner, R. Reynolds and G. F. Sensabaugh

University of California
Berkeley, California

Restriction fragment length polymorphism (RFLP) analysis can be done only on samples containing a minimum quantity of high molecular weight DNA. Samples collected from deceased individuals, however, often contain degraded DNA. It is important to know, therefore, whether some postmortem tissues are better than others with regard to the quantity and quality of recoverable DNA. In this study, 7 tissues from 20 cadavers were evaluated for DNA recovery.

MATERIALS AND METHODS

The cadavers were selected at random and as such vary with regard to age, sex, cause of death, and time from death to sample collection. The 7 tissues collected were heart, spleen, liver, muscle, blood, hair and bone. The heart, spleen, liver and muscle samples were prepared by grinding in a mortar and pestle under liquid nitrogen. The nucleated cells from the previously frozen blood samples were collected by centrifugation. The hair root was removed from the shaft. All samples were digested with proteinase-K followed by phenol/chloroform extraction to isolate the DNA fraction. The DNA fraction was concentrated by ethanol precipitation in most cases; hair root DNA was concentrated by spin dialysis using Centricon-30 units. The DNA quantity and purity was determined by OD 260/280 readings. The DNA quality (molecular weight) was determined by agarose gel electrophoresis followed by ethidium bromide visualization.

RESULTS

Although high molecular weight DNA was obtained from each type of tissue on occasion, no 1 tissue was predictable and reproducible as a reasonable source of large quantities of high molecular weight DNA. Moreover, the time interval between death and sample collection was not a reliable indicator of DNA status; some short postmortem samples contained badly degraded DNA, whereas some long interval samples contained good quality DNA. Almost every tissue type yielded DNA that was a mixture of various size fragments. Blood, heart and bone normally had good quality DNA but not in every instance. Hair roots generally had high molecular weight DNA but not in large quantity. The muscle samples often yielded large amounts of badly degraded DNA (less than 500 bp). Spleen samples were unpredictable as to quality of DNA but were easy to grind. The liver samples rarely yielded high molecular weight DNA.

The results of this survey suggest that successful DNA typing of cadavers may require the collection of samples of several tissues including blood, heart, bone, and numerous strands of plucked hair.

BLANK PAGE

CONSEQUENCES OF NUCLEOTIDE MISINCORPORATION DURING THE POLYMERASE CHAIN REACTION

G. F. Sensabaugh

University of California
Berkeley, California

The polymerase chain reaction (PCR) allows the selective replication of short segments of DNA sequence from target genomes; as many as 107 or more copies of a target sequence can be produced by this process. The fidelity of the replication process is a relevant forensic concern. If PCR yields true copies, that is, copies that genetically type true, then the concern is without foundation. If, however, the level of nucleotide misincorporation is high enough to produce a detectable number of copies showing a different genetic type than that of the template, then PCR would be unacceptable for forensic use.

MATERIALS AND METHODS

It is well recognized that DNA polymerases occasionally make mistakes. For Taq polymerase, the polymerase most commonly employed in PCR, the misincorporation rate has been independently estimated at 2×10^{-4} (Saiki *et al.* 1988) and at 1.1×10^{-4} (Tindall and Kunkel 1988). This presentation characterizes the PCR products resulting from misincorporation assuming the higher rate, that is, the 2×10^{-4} rate.

A base substitution generated by a misincorporation event will be propagated in all the progeny of the DNA strand in which the misincorporation occurs. Accordingly, the proportion of the total PCR product resulting from any single misincorporation event depends on the initial number of template sequences n and the cycle r at which the misincorporation event occurs. The general equation for this proportion is $(1/n)(\frac{1}{2})^{r+1}$; the proportion values for various initial template numbers and cycle numbers are shown in Table 1. It is clear that an incorrect sequence resulting from any single misincorporation event would be detectable only if the initial copy number is less than 10 and the misincorporation occurs at an early cycle.

Table 1. PROPORTION OF PCR PRODUCT STRANDS CONTAINING AN AMPLIFICATION ERROR

Initial Copies DNA	Cycle of Error Substitution			
	1	2	5	r
1	.25	.125	.015	$\frac{1}{2}^{r+1}$
10	.025	.0125	.0015	$.\frac{1}{2}^{r+1}$
100	.0025	.00125	.00015	$.0\frac{1}{2}^{r+1}$

RESULTS

The proportion of copies suffering misincorporation hits in each cycle depends on the misincorporation rate and the length of the sequence being replicated. At the end of any cycle, the proportion of PCR products containing no misincorporations, 1 misincorporation anywhere in the sequence, a misincorporation at the ith position in the sequence, and 2 or more misincorporations anywhere are defined by a Poisson distribution; some representative values are shown in Table 2. It is evident that the proportion of PCR products containing any specific misincorporation, that is, a base difference that might change one genetic type to another, would be undetectable.

Table 2. DISTRIBUTION OF PCR PRODUCTS AT THE END OF A CYCLE CONTAINING 0, 1, AND 2 MISINCORPORATIONS

Sequence Length	Single Base Misincorporation			Two or More
	None	At base i	All	
100 bp	0.9802	0.000196	0.0196	0.0002
250 bp	0.9512	0.000190	0.0476	0.0012
500 bp	0.9048	0.000181	0.0905	0.0047
1000 bp	0.8187	0.000164	0.1638	0.0175

In conclusion, although a significant proportion of PCR products contain some misincorporation, the proportion containing any single base substitution is below the level of detection, and hence would not result in typing error.

REFERENCES

Tindall, K. R. and Kunkel, T. A. (1988). Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase, Biochem. 27:6008–6013.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, Science 239:457–491.

264

DNA IN HAIR

C. H. Von Beroldingen, R. K. Roby and G. F. Sensabaugh

University of California
Berkeley, California

S. Walsh

Department of Human Genetics, Cetus Corporation
Emeryville, California

The ability to carry out DNA typing on single human hairs can be of considerable forensic importance. The quantity and quality of DNA present in a hair determines the analytical approach (that is, RFLP analysis or PCR analysis) and the success rate. We have investigated the amount and type of DNA present in plucked hairs, shed hairs and hair shafts and evaluated the success rates for DNA typing from single hairs using PCR analysis.

MATERIALS AND METHODS

The DNA was isolated from the root portion of single hairs as described by Higuchi *et al.* (1988). The DNA was isolated from 0.5 g of cut hair shafts as above except that the digestion took place in a volume of 25 ml. For the direct lysis procedure, hairs were digested in the presence of 0.45% NP-40/Tween rather than SDS. The DNA content of hairs was measured fluorometrically by a modification of the procedure of Hukkelhoven *et al.* (1981). Procedures used for HLA DQα amplification and typing were as described by Higuchi *et al.* (1988). In some cases, typing was performed by hybridizing biotinylated PCR product with immobilized allele-specific oligonucleotide probes (Saiki *et al.* 1989).

RESULTS

The DNA from over 200 single hairs, both plucked and shed, has been amplified by the polymerase chain reaction (PCR) and tested for genetic type at the DQα locus. The genetic types of the hair donors have been determined independently by testing on blood samples. In general, DNA extracted from plucked hairs yields higher levels of PCR product than DNA from shed hair and is more likely to give unambiguous typing results. Amplification of DNA from shed hairs is more susceptible to problems associated with contamination. When steps are taken to avoid contamination, one may expect to find a type greater than 90% of plucked hairs and up to 50% of shed hairs.

One method to minimize the introduction of adventitious contamination during sample preparation is to eliminate DNA extraction steps by adding an aliquot of the hair lysate directly to the amplification reaction. Preliminary results indicate that about 40% of shed hairs may be amplified and typed using this procedure.

To further investigate the state of DNA in hair, we have isolated and characterized DNA from individual plucked hair roots, pooled shed hairs and pooled hair shafts. The DNA was assessed for quality by gel electrophoreses. Plucked single hairs retaining sheath material contain mostly high MW DNA. Individual shed hairs usually contain too little DNA to allow assessment of quality. However, by pooling groups of shed hairs, high MW DNA may be visualized on an agarose gel. Concentration of DNA isolated from pooled hair shafts revealed the presence of some high MW DNA as well as substantial quantities of low MW nucleic acid. DNA content was measured fluorometrically in single plucked hairs and pooled shed hairs and estimated by a calibration gel for pooled hair shafts. These results, summarized in Table 1, indicate that single plucked hairs may be suitable for RFLP analysis, whereas DNA from single shed hairs or hair shafts could only be analyzed by PCR.

Table 1. DNA IN HAIR

Type	Number	Average	Range
PLUCKED (with sheath)	36	375 ± 216 (ng/hair)	1–784 (ng/hair)
PLUCKED (without sheath)	2	54 (ng/hair)	43, 64 (ng/hair)
SHED	43	3 (ng/hair)	0.8–12 (ng/hair)
SHAFT	6 (individuals)	9 (pg/cm)	0.2–40 (pg/cm)

REFERENCES

Higuchi, R., Von Beroldingen, C. H., Sensabaugh, G. F. and Erlich, H. A. (1988). DNA typing from single hairs, Nature 332:543-546.

Hukkelhoven, M. W. A. C., Vromans, E., Markslag, A. M. G. and Vermorkin, A. J. M. (1981). A simple fluorimetric microassay for DNA in hair follicles or fractions of hair follicles, Anticancer Res. 1:341-344.

Saiki, R. S., Walsh, P. S., Levenson, C. H. and Erlich, H. A. (1989). Genetic analysis of amplified DNA with immobilized sequence specific oligonucleotide probes, Proc. Natl. Acad. Sci. USA (in press).

DNA AMPLIFICATION AND TYPING FROM AGED BIOLOGICAL EVIDENCE

E. T. Blake

Forensic Science Associates
Richmond, California

S. Paabo

Department of Biochemistry, University of California
Berkeley, California

M. D. Stolorow

Cellmark Diagnostics
Germantown, Maryland

We describe 2 cases in which DNA was successfully typed from aged and degraded evidence specimens. These cases illustrate the capacity of the polymerase chain reaction (PCR) to amplify genetically informative DNA from the most difficult forensic specimens.

The first case is the Illinois v. Gary Dotson case. Dotson was convicted of the 1977 alleged rape of Cathleen Crowell Webb. In 1985 Webb recanted her testimony, asserting that the alleged rape never occurred and explaining the presence of semen in her panties as the result of consensual intercourse with her boyfriend. Webb's recantation was not adequate to reverse Dotson's conviction. In 1987 Alec Jeffreys unsuccessfully attempted an RFLP analysis on the semen stain from the panties. Following the Jeffreys' tests, the panties were submitted to the Illinois State Police Laboratory for PCR employed DQα typing. Mr. Stolorow, then of the State Police Laboratory, monitored the work. The testing's purpose was to determine whether or not Gary Dotson and/or Webb's boyfriend, David Beirne, could be eliminated as potential sperm donor(s) on the Webb panties.

MATERIALS AND METHODS

The original panties stain was approximately 2 x 8 inches and extended from the crotch to the rear waist band. Previous examiners sampled this stain extensively. Two specimens for testing were removed from near the panties crotch (area YY) and near the rear waist band (area ZZ). Cellular debris was extracted from the fabric by immersing (1 cm²) cuttings in a small test tube containing 1 ml of H_2O for at least 1 hr., occasionally scrubbing the fabric with a sterile toothpick. Each piece of fabric was extracted twice, and an aliquot (5%) of the debris from each area was examined microscopically. The cellular debris was differently, yet essentially, digested as recommended by Gill *et al.* (1985). The epithelial cells were preferentially digested by adding 0.5 ml digest buffer (10 mM Tris, 10 mM EDTA, 50 mM NaCl, 2% SDS, pH 7.5) and 15 μl of proteinase K stock (10 mμ/ml) followed by incubation at 60°C for 1 hr. Following incubation, the sperm were made into pellets, the supernatant containing epithelial cell DNA was removed and saved, and the sperm were washed 3 times with digest buffer and once with sterile water. An aliquot (5–10%) of the sperm pellet was removed for microscopic inspection to monitor the success of the epithelial cell digestion. The sperm pellet was digested by addition of 0.5 ml digest buffer, 20 μl of 1 M DTT, and 20 μl proteinase K followed by incubation at 60°C over night. The digested DNAs were extracted with chloroform/phenol (50:50) followed by n-butanol. The extracted DNAs were washed 3 times in TE (10mM Tris, 0.1 mM EDTA) and concentrated to 25–40 μl using centricon 30 molecular filters (Amicon). Concentrated DNAs were evaluated using minisubmarine electrophoresis gels employing 1% agarose (Seaken, GTG) in TBE buffer (0.089 M Tris, 0.089 M Boric Acid, 2.5 mM EDTA, pH 8.3, and containing ethidium bromide at 0.5 μg/ml) against a reference DNA ladder consisting of 1 μg DNA restricted with *Hind*III. The DQα gene from the concentrated DNA was amplified essentially as described by Higuchi *et al.* (1988) and Bugawan *et al.* (1988) for 40 cycles of denaturation (30 sec. at 94°C), annealing (30 sec. at 55°C), and extension (30 sec. at 72°C) employing PCR cocktails and thermal cycler provided by Cetus (Emeryville, CA). The PCR product was evaluated in minisubmarine electrophoresis gels employing the procedure described previously. However, the agarose was composed of

1% Seakem GTG and 3% Seakem NuSieve, and the reference DNA was 1 μg X174 DNA restricted with HaeIII or pBR322 DNA restricted with HaeIII. The amplified DQ genes were typed using a sot blot format according to Bugawan's *et al.* (1988) procedure and subsequently using an immobilized allele specific probe format (Saiki *et al.* 1989).

RESULTS

The DQα typing could distinguish between Webb (1.1, 3), Dotson (1.1, 4) and Beirne (2, 3). Furthermore, Dotson and Beirne do not share any alleles; thus, a mixture of sperm from these 2 individuals could be recognized potentially. The panties stain analysis revealed that the sperm donor was DQα type 2, 3. This type was also obtained from DNA prepared by Jeffreys in England and amplified and typed in California. The DQα type 2, 3 occurs in approximately 5% of the Caucasian population. The data demonstrate that the sperm in the Webb panties did not originate from Dotson and are compatible with originating from Webb's boyfriend, Beirne, and 5% of the Caucasian population.

CASE #2

The second case involved a female child who was discovered missing from her parents' desert community in San Bernardino County, California. Approximately 18 months later the skull cap from a human child was found in the desert, 2 miles from the home of the missing child. At issue is whether or not the found skull cap represents the remains of the missing child. We have approached this problem using genetic reconstruction from the missing child's parents. This reconstruction is based on the nuclear DQ gene and the maternally inherited mitochondrial genome. The latter work is still in progress.

MATERIALS AND METHODS

The DNA from the bone was prepared 2 ways. First, bone (½ cm²) was chipped from the central portion of the skull. This bone chip was di-gested without further preparation in 0.5 ml digest buffer, 20 μl of 1 M DTT, and 20 μl of proteinase K stock at 60°C. A second aliquot of DTT and proteinase K was added after 3 hr. of initial digestion. In the second procedure, bone of similar size was crushed prior to the digestion described above. The digestion process did not soften or otherwise alter the bone's structural integrity. The digest fluid was extracted, concentrated and evaluated, and the DQ gene was amplified and typed as described above. Material scraped from the skull's surface served as a negative control

RESULTS

The DNA recovered from the bone was highly degraded varying in size from 5 Kb to 100 bp. The amount of DNA recovered from the crushed bone was 50–100 times greater than the DNA recovered from the uncrushed bone fragment. Thus, it would appear that DNA recovery is a function of freshly exposed bone surface area. The DQα gene amplification was successful from both bone preparations. The bone donor's DQ type was determined to be type 4, 4. By using DQα typing alone, these 2 types together occur as the parents' offspring in approximately 18% of Caucasians.

REFERENCES

Bugawan, T. L., Saiki, R. K., Levenson, C. H., Watson, R. M. and Erlich, H. A. (1988). The use of non-radioactive oligonucleotide probes to analyze enzymatically amplified DNA for prenatal diagnosis and forensic HLA typing, Bio/Technology, 6:943–947.

Gill, P., Jeffreys, A. J. and Werrett, D. J. (1985). Forensic application of DNA fingerprints, Nature, 318:577–579.

Higuchi, R., von Beroldingen, C. H., Sesabaugh, G. F. and Erlich, H. A. (1988). DNA typing from single hairs. Nature, 332:543–546.

Saiki, R. K., Walsh, P. S., Levenson, C. H. and Erlich, H. A. (1989). Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes, PNAS, in press.

THE EFFECT OF DETERGENTS ON DNA IN BLOOD AND SEMEN

H. C. Lee and E. M. Pagliaro

Connecticut State Police Forensic Science Laboratory
Meridian, Connecticuit

R. E. Gaensslen

University of New Haven Forensic Sciences Program
West Haven, Connecticuit

T. Keith and S. Rose

Collaborative Research, Inc.
Bedford, Massachusetts

The introduction of DNA analysis into forensic science casework is revolutionizing the traditional procedures used in examining biological evidence, and reliable DNA typing methods have been developed (Jeffreys *et al.* 1985a,b; Gill 1985; Giusti *et al.* 1986; Kanter *et al.* 1986; Budowle *et al.* 1989). With them, unequivocal results can be obtained with forensic samples if well established procedures and proper controls are used. Some studies have been conducted on the effect of various potential environmental influences and contaminants on DNA (Adams 1988; Adams *et al.* 1989), but no detailed investigation has been carried out on the effects of common household cleansing and personal hygiene agents on DNA and DNA typing. Recently, several homicide cases occurred in which attempts were made to use detergents in cleaning up bloodstains. Several sexual assault cases have also arisen in which the victim was forced to wash and/or douche after the assault. Samples from these cases were subjected to DNA analysis, but their contamination with the commercial products apparently caused inconclusive results. To understand the effects of common cleansers and personal hygiene products on DNA in blood and semen, a series of experiments has been conducted.

MATERIALS AND METHODS

Blood samples collected from 10 normal, healthy donors in EDTA were used to make bloodstains on clean cotton cloth that were air dried at room temperature. Blood samples were treated with various detergents as follows: (1) Cleansing agent mixed with liquid blood at concentrations of 1, 10, 25 and 50% (v/v or w/v), stains prepared from the mixtures and air dried at room temperature; (2) Cleansing agent mixed 1:1 with blood and stains prepared from the mixtures subjected to different cleaning manipulations.

The DNA was isolated by cutting up a 2 cm$_2$ portion of stain, digesting with proteinase K, standard phenol-chloroform extraction, and ethanol precipitation (Maniatis *et al.* 1982). The DNA was subjected to electrophoresis on a yield gel to estimate the concentration of high MW DNA and the degree of degradation. The DNA from some specimens was digested with *Hae*III, subjected electrophoresis in an agarose typing gel, transferred to a nylon membrane by Southern blotting (Maniatis *et al.* 1982; Southern 1975), and hybridized with human DNA probe CRI-PAT L1 1077-1 following Collaborative Research standard protocols. The RFLP patterns were detected autoradiographically.

Fresh semen obtained from 6 normal individuals was mixed with douche liquid, soap, contraceptive sponge liquid and stains prepared from the mixtures (and control semen) on clean cotton cloth. The DNA extraction, evaluation, digestion and RFLP analysis were carried out by standard procedures as previously described.

Table 1. HOUSEHOLD CLEANSERS USED IN TREATING BLOODSTAINS

Laundry Products
  Era
  Cheer (powder)
  Clorox bleach
  Sure-fine all fabric bleach
  All
  Woolite
  Tide (powder)
  Generic "all purpose" detergent
  Shout
Shampoos
  Awaphui
  Prell

Neutrogena T
Systema Sebastian
Jermack
Selsun Blue
Clean & Clear
Carpet Cleaners
Glory
Woolite
Soaps
Dial antibacterial liquid
Ivory liquid
Dove
Dishwashing Products
Dawn
Palmolive liquid

Table 1. HOUSEHOLD CLEANSERS USED IN TREATING BLOODSTAINS—Continued

Palmolive dishwasher liquid
Generic dishwashing liquid
All Purpose Cleaners
"Gleme" glass cleaner
Pine-sol
Lysol Tub & Tile Cleaner
Soft Scrub
Bab-O
Lysol Disinfectant Cleaner
Mr. Clean
Spic and Span (powder)
Fantastik
Ammonia
Murphy's Oil Soap
Glass Plus

Table 2. APPROXIMATE DNA YIELD (NG) OF HMW DNA FROM BLOODSTAINS TREATED WITH INDICATED CLEANSER AT INDICATED CONCENTRATION

Cleanser	Concentration (%)			
	1	10	25	50
Era	700	700	600	1000
Cheer	900	800	650	200
Bleach	1000	700	850	800
All Fabric bleach	700	500	600	600
Awapuhi	700	850	700	500
Prell	500	700	500	700
Dawn	1500	1000	800	750
Palmolive dishwasher liquid	250	300	250	125
Ivory	850	800	850	700
Mr. Clean	900	750	800	750
Spic and Span	700	600	500	500
Fantastik	1000	900	1000	850

RESULTS

Table 1 shows the cleansing agents used in this study. Table 2 shows the average yield of high MW DNA from blood and cleansing agent mixtures at various concentrations. Figure 1 shows representative data obtained by yield gel analysis of DNA from bloodstains subjected to various cleansing concentrations. Figure 1 shows representative data obtained by yield gel analysis of DNA from bloodstains subjected to various cleansing modalities. Samples extracted and analyzed after one month demonstrated similar results. Sufficient HMW DNA for RFLP analysis was obtained in all cases (Table 2), although there was some variation. The DNA yield from scrubbed, cleanser-treated specimens was less in most cases than the 800–1500 ng obtained from control samples. Some degradation of DNA was noted with some cleansing agents (Figure 1). However, no significant alteration of DNA quality was observed in most cases, and DNA of sufficient quantity for RFLP analysis was obtained. Figure 2 shows an autoradiogram of RFLP patterns from treated and untreated samples. Treatment of blood with cleansing agents did not cause any alteration in the RFLP pattern. Semen samples treated with douche, soap, and contraceptive yielded DNA of quantity and quality comparable to that of control specimens.

As shown, blood and semen samples treated under the conditions described with a variety of common household and personal hygiene cleansing products will yield high MW DNA usually comparable in quantity to control specimens. Some degradation in detergent treated samples will occur. Sufficient high MW DNA was obtained from all samples tested for RFLP typing. No band shifts or other alterations in RFLP patterns were noted in treated specimens in comparison with controls.
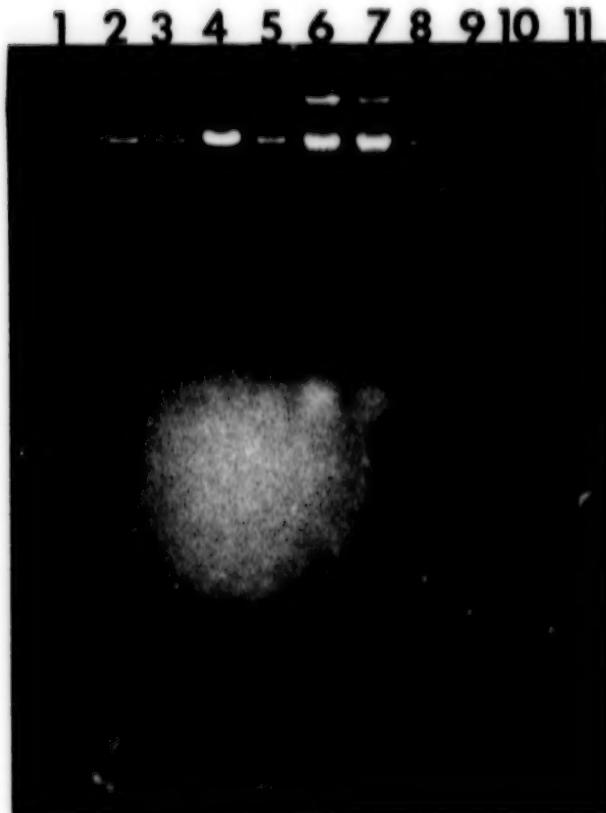
figure 1

Figure 1. Representative Yield Gel. Lanes 1-4 contain 17, 58, 116 and 174 ng DNA, respectively, as concentration markers. Lanes 5 to 16 contain DNA from bloodstains made from blood mixed with various cleansing agents and scrubbed.

figure 2

Figure 2. Autoradiogram showing HaeIII Restriction Fragments of DNA from Bloodstains Treated with Various Detergents and Control Bloodstain following Hybridization with CRI-PAT L1 1077-1.

REFERENCES

Adams, D. E. (1988). Validation of the FBI procedure for DNA analysis: a summary, Crime Lab Digest 15:106-108.

Adams, D. E., Baechtel, F. S. and Budowle, B. (1989). DNA analysis of body fluid stains subjected to environmental, chemical and biological insults, presented at 41st Annual Meeting, Am. Acad. Forensic Sci., Las Vegas, NV.

Budowle, B., Deadman, H., Murch, R. and Baechtel, F. S. (1989). An introduction to the methods of DNA analysis under investigation in the FBI Laboratory, Crime Lab Digest 15:8-21.

Gill, P., Jeffreys, A. J. and Werrett, D. J. (1985). Forensic application of DNA fingerprints, Nature 318:577-579.

Giusti, A., Baird, M., Pasquale, S., Balasz, I. and Glassberg, J. (1986). Application of Deoxyribonucleic Acid (DNA) polymorphism to the analysis of DNA recovered from sperm, J. Forensic Sci. 31:409-417.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985a). Hypervariable minisatellite regions in human DNA, Nature 314:67-73.

Jeffreys, A. J., Wilson, V. and Thein, S. L. (1985b). Individual specific fingerprints of human DNA, Nature 316:76-79.

Kanter, E., Baird, M., Shaler, R. and Balasz, I. (1986). Analysis of restriction fragment length polymorphism in deoxyribonucleic acid (DNA) recovered from dried bloodstains, J. Forensic Sci. 31:403-408.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis, J. Molec. Biol. 98:503-517.

BLANK PAGE

A STRATIFIED APPROACH TO FORENSIC IDENTIFICATION BY DNA ANALYSIS: A COMPARISON OF TRADITIONAL GENETIC SYSTEMS AND RESTRICTION FRAGMENT LENGTH POLYMORPHISMS (RFLPs)

R. T. Acton, T. W. Hodge, R. C. P. Go, S. C. Peiper and L. Harman

University of Alabama at Birmingham
Birmingham, Alabama

Recombinant DNA technology has increased the power to identify individuals involved in paternity disputes and forensic testing. Two cases have been selected for presentation to illustrate the power of traditional genetic analysis systems compared to analysis at the DNA level for the purpose of establishing individual identities.

Case 1 involved a 13 year old female who arrived at UAB Hospital in labor and delivered a 36-week stillborn anencephalic female fetus. It was suspected that the girl had been sexually abused by her father and that the fetus was the product of this liason. A court order directed our laboratory to determine the probability that the girl's father was also the father of the stillborn fetus. Seven days had elapsed since the birth of the fetus and the only biological material available was about 2 ml of blood that had been maintained at about 5°C and paraffin blocks of formalin-fixed tissue. We were able to obtain a sufficient quantity of material from the blood of the fetus to conduct the genetic tests shown in Table 1. Based on the 11 traditional genetic systems utilized, the probability that the putative father (PF) was the biological father of the fetus was calculated to be 99.40%. An RFLP analysis was performed to further document the relationship between the fetus and the PF. Southern (1975) blots containing PvuII digested DNA from the mother, fetus and PF were annealed with radiolabelled pDP001, a multilocus probe that detects class 1 genes within the major histocompatibility complex (MHC) of chromosome 6 (Sood *et al.* 1981). Although the PF was not excluded by use of this probe, the probability of paternity based on these studies alone was low due to the low frequencies of RFLPs shared by the fetus and PF but not by the mother. This condition is not unexpected since the mother, her natural father, also the PF, and the fetus are identical at 3 of the HLA-A & B loci. We attempted to extract DNA from formaldehyde fixed fetal tissues for evaluation by PCR and dot blot analysis using MHC class II probes. The results of these studies have been equivocal, at least in part due to the small amount of DNA that can be extracted from 8μm sections. The kindred was also analyzed by use of probe pYNH24 (GenMark, Salt Lake City, UT) that is derived from a locus on chromosome 2 that has a variable number of tandem repeats (VNTR) (Nakamura *et al.* 1987). We have observed pYNH24 VNTR HaeIII fragments ranging from 0.7 Kb to 6.8 Kb in our local white population. These fragments were arbitrarily assigned to bins which differ by 100 bp. This method allowed us to define 34 alleles. It is predicted that 82.7% of wrongly identified white persons can be excluded with this probe alone. In case 1 the probability of paternity based on the use of the pYNH24 probe alone was 98.51%.

Table 1. COMPARISON OF TRADITIONAL GENETIC SYSTEMS AND DNA ANALYSIS IN PATERNITY TESTING OF FETAL MATERIAL

Genetic System	Mother	Fetus	Putative Father	Paternity Index
HLA-A	A2,A3	A2,A3	A2,A3	
HLA-B	B7,B27	B7,B27	B18,B27	6.66
ABO	A	A	A	1.32
Duffy	Fy(a+b+)	Fy(a+b−)	Fy(a+b−)	2.26
Kell	k	k	k	1.05
Kidd	Jk(a+b−)	Jk(a+b−)	Jk(a+b−)	1.94
MNS	MNs	MNs	Ns	1.47
Rh	cDe	cDe	CDe	1.24
EsD	1-2	1-2	2	1.00
PGM1	1+	1+	1+	1.57
Pi	M1M2	M1M2	M1M2	1.17
TF	C1	C1	C1	1.25
pYNH24/HaeIII	1.6/1.1	1.6/1.1	2.2/1.1	65.8

Traditional Genetic Systems Cumulative Paternity Index: 168 to 1
Traditional Genetic Systems and RFLP Cumulative Paternity Index: 11054 to 1
Traditional Genetic Systems Probability of Paternity: 99.40%
RFLP Probability of Paternity: 98.51%
Traditional Genetic Systems and DNA Probability of Paternity Index: 99.99%

Case 2 consisted of a married woman who was raped. The victim subsequently learned she was pregnant with twins. The woman was referred to the Laboratory of Medical Genetics at UAB where an amniocentesis was performed to ascertain whether the father of the twins was her husband since the rapist was not available for testing. Amniotic cells were obtained from the sacs of both twins at about 14 weeks of gestation and cord blood from one of them at about 18 weeks of gesta-

tion and referred to our laboratory. As shown in Table 2, the cord blood from fetus 1 was assayed for the presence of HLA and red blood cell antigen polymorphisms. The probability that the woman's husband was the father of fetus 1 was determined to be 99.76%. Based on assessment of the amniotic cells from both fetuses by the pYNH24 probe the probability that the woman's husband was the biological father of fetus 1 and fetus 2 was determined to be 97.04% and 98.50%, respectively.

Table 2. COMPARISON OF TRADITIONAL GENETIC SYSTEMS AND DNA ANALYSIS IN PATERNITY TESTING OF FETAL AMNIOTIC AND CORD BLOOD CELLS

Genetic Systems	Mother	Fetus 1	Fetus 2	Putative Father	Paternity Index Fetus 1	Paternity Index Fetus 2
HLA-A	A1,A2	A2	....	A2	....	....
HLA-B	B7,B8	B7,Bw62	....	Bw62	30.09	....
ABO	A	A	....	0	0.91	....
Duffy	Fy(a−b+)	Fy(a+b+)	....	Fy(a+b+)	1.18	....
Kell	k	k	....	k	1.05	....
Kidd	Jk(a+b−)	Jk(a+b−)	....	Jk(a+b−)	1.94	....
MNS	MNSs	MNS	....	MNS	5.36	....
Rh	cDE	CDE	....	CDe	1.19	....
pYNH24/HaeIII	2.6/1.6	2.4/1.6	2.7/1.6	2.7/2.4	32.89	65.79

Fetus 1 Traditional Genetic Systems Cumulative Paternity Index: 420 to 1
Fetus 1 Traditional Genetic Systems and RFLP Cumulative Paternity Index: 13814 to 1
Fetus 1 Traditional Genetic Systems Probability of Paternity: 99.76%
Fetus 1 RFLP Probability of Paternity: 97.04%
Fetus 1 Traditional Genetic Systems and RFLP Probability of Paternity: 99.99%
Fetus 2 RFLP Probability of Paternity: 98.50%

In addition to the pYNH24 probe we have evaluated another VNTR locus flanking the insulin gene on chromosome 11 by probe phins 310 which

detects PvuII alleles that range in size from 0.5 Kb-5 Kb. The probability of excluding a wrongly identified white individual by use of this probe is

74.9%. This probe did not exclude either PF in cases 1 or 2. The use of pYNH24 and phins 310 together in a given case provides the power to exclude 95.65% of wrongly identified white individuals.

In summary, probes detecting VNTR loci provide powerful genetic systems for identity purposes. Due to the difficulties in amplifying sufficient quantities of DNA extracted from formalin fixed, paraffin-embedded tissue we would suggest that medical examiners be advised to maintain tissues of victims in 50% ethanol. High molecular weight DNA can be extracted from such specimens up to 6 years after fixation (Smith *et al.* 1987). The availability of probes detecting highly polymorphic loci and ability to analyze DNA extracted from tissues held over varying lengths of time enhances paternity testing and forensic examinations.

REFERENCES

Bell, G. I., Karam, J. H. and Rutter, W. J. (1981). Polymorphic DNA region adjacent to the 5' end of the human insulin gene, Proc. Natl. Acad. Sci. USA 78:5759–5763.

Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Mark, H., Kumlin, E. and White, R. (1987). Variable number of tandem repeat (VNTR) markers for human gene mapping, Science 235:1616–1622.

Smith, L. J., Braylan, R. C., Nutkis, J. E., Edmundson, K. B., Downing, J. R. and Wakeland, E.K. (1987). Extraction of cellular DNA from human cells and tissues fixed in ethanol, Anal. Biochem. 160:135–138.

Sood, A. K., Pereira, D. and Weissman, S. M. (1981). Isolation and partial nucleotide sequence of a cDNA clone for human histocompatibility antigens HLA–B by use of an oligodeoxynucleotide primer, Proc. Natl. Acad. Sci. USA 78:616–620.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517.

BLANK PAGE

APPLICATION OF A NEW DNA PROBE TO FORENSIC ANALYSIS

H. C. Coleman

GeneLex Corporation
Seattle, Washington

D. C. MacClaren

Washington State Patrol
Seattle, Washington

K. W. van Dijk, C. J. Lotshaw and E. C. B. Milner

Virginia Mason Research Center
Seattle, Washington

An extremely sensitive, highly polymorphic variable number tandem repeat (VNTR) human DNA clone, pH30, has been isolated and found to be useful for the typing of bio-forensic evidence (Milner *et al.* 1989).

Gene mapping studies with multiple mouse/human somatic cell hybrids were used to localize pH30 to chromosome 4. This locus has been assigned the Human Gene Mapping Workshop number D4S139 (Figure 1). Mapping was accomplished by linkage analysis performed on 18 large 3 generation and six 2 generation families (CEPH families) using 53 chromosome 4 markers. Linkage was found at a distance of 16 cm, and a LOD score greater than 3, to L231 and C82, markers previously mapped near the telomere of the chromosome 4 long arm. Heterozygosity was observed at 89% in Caucasoids. Over 50 bands have been detected by pH30, the most common occurring at a frequency of 13 to 15% depending on the ethnic/racial composition of the population group under study (unpublished). Clone pH 30 is complimentary to the VNTR sequence as it detects restriction fragment variants in human DNA with over 20 different restriction enzymes. There is also a rare variant that has a Hae III site within the VNTR locus. Plasmid pH30 detects relatively large restriction fragments on Southern blots of human DNA, (avg. 9 Kb) when compared to most other VNTR probes (avg. 3 to 4 Kb). This characteristic makes ph30 a highly useful forensic DNA Typing analysis.

The DNA probe pH30 has been used on Southern blots in a series of nonprobative sexual assault evidence swabs. The swabs were stored frozen at $-20\,^{\circ}$C for periods of up to 1 year prior to analysis. Figure 2 illustrates typical patterns obtained when the blots are first probed with pH30 and then a Y chromosome specific probe. The pH30 demonstrates a high level sensitivity, being able to detect single copy VNTR sequences in approximately 25 ng of DNA (compare case 2, lane v with 0.1 μg standard lane). Evidence of sexual intercourse is illustrated in 2 cases which were positive for Y chromosomes. The vaginal swab in case 1 contained 2 sets of bands, 1 from the victim, and 1 from the presumed assailant. In case 2 the victim's DNA was present on the vaginal swab (Y chromosome negative) while the presumed assailant's DNA was present on the rectal swab (Y chromosome positive). In case 2 the victim and the assailant share the lower molecular weight band which illustrates the importance of sequential probing.

The D4S139 VNTR bands were detected in 93% of cases (n=43) (Table 1), and 85% of all swabs (n=118) contained DNA. Of these, 73% (n=89) had readable VNTR patterns. Thirty percent of the female case derived swabs (n=33) had Y chromosome specific DNA present, indicating that a Y chromosome was present. Eighty-nine percent of oral swabs (n=28) contained DNA, suggesting that buccal swabs, saliva or dried saliva stains may be a useful source of DNA.

Observed data on the VNTR probe pH30 suggest that it is highly useful for DNA typing of bioforensic evidence. It exhibits a high level of polymorphism and heterozygosity and may be the most sensitive probe described thus far.

REFERENCES

Milner, E. C. B., Lotshaw, C. L., van Dijk, K. W., Charmley, P., Concannon, P. and Schroeder, H. W., Jr. (1989). Isolation and mapping of a polymorphic DNA sequence pH30 on chro-

mosome 4 [HGM provisional no. D4S139], Nucl. Acids Res. 17:4003.

Table 1. ANALYSIS OF DISCARDED SEXUAL ASSUALT EVIDENCE SWABS.

	DNA +	Y Cx +	VNTR + (pH30)	P30 +
CASE TOTALS	93% 40/43	30% 10/33	93% 27/29	43% 6/14
VAGINAL				
SWABS	94% 50/53	23% 11/47	94% 35/37	38% 6/16
ORAL SWABS	89% 25/28	....	74% 17/23	20% 1/5
RECTAL SWABS	92% 11/12	17% 2/12	64% 7/11	....
TOTAL SWABS	85% 100/118	19% 17/89	73% 65/89	31% 9/29

DNA was isolated and probes as in Figure 2.
DNA +, DNA detectable by ethidium bromide stained minigel
Y chromosome +, Y chromosome specific sequences present
VNTR +, Detectable band pattern by DNA probe pH30
P30 +, Seminal vesicle protein by crossed immunoelectrophoresis.

MAP POSITION OF pH30
(D4S139)

Chromosome 4
Long arm
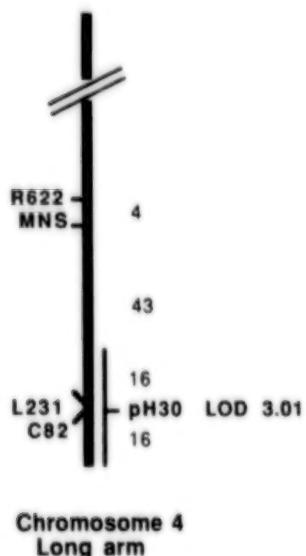
Figure. 1. Map of the telomeric region of the long arm of Chromosome 4 showing the mapped position of pH30/D4S139. The polymorphic DNA sequence pH30 was isolated and mapped (Milner et al. 1989). The bar labeled pH30 indicates that pH30 has been mapped at least 16 cM from the L231/C82 region on chromosome 4. L231, C82 and R622 are DNA restriction site polymorphisms, and MNS is the blood group antigens. LOD is log of the odds. The telomere is at the bottom of the diagram. Un-bolded numbers on the right side of the diagram indicated map distances in centi-Morgans (cM).

VNTR Probe

Y Chromosome Probe
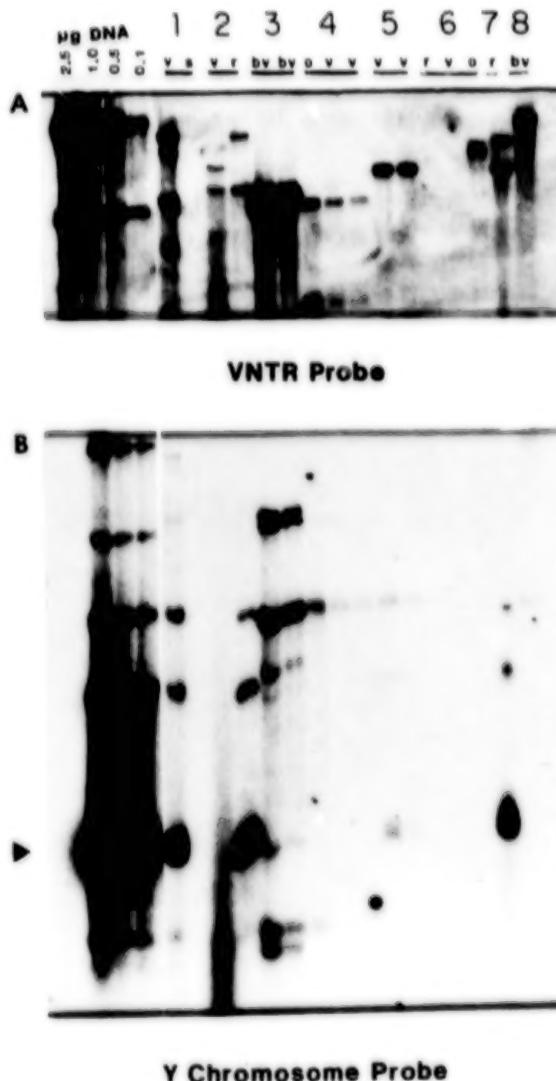
Figure. 2. Southern blot of Hinf1 digested DNA obtained from discarded sexual assault evidence swabs. DNA was eluted from sexual assault evidence swabs (previously stored at −20 °C) by extraction in 300 ml 100 mM NaCl, 25 mM EDTA, 10 mM Tris pH 8 for 30 min. at 25 °C, with mixing vigorously every 10 min. The 10 ml aliquots were withdrawn for acid phosphatase and P-30 analysis. The 30 ml of 10% SDS were added and the incubation continued, as above, for another 30 min. Eluate was centrifuged away from swabs and brought to 200 μg/ml Proteinase K, 39 mM dithiothreitol and incubated at 55 °C 1 hr. to overnight. The DNA was purified by phenol chloroform extraction and ethanol precipitation prior to digestion with HinF I, and Southern blotting. The blot was hybridized to pH30 (Panel A) and then to a Y chromosome specific probe (Amersham) (Panel B) without intermediate stripping of the blot. Hybridized blots were exposed overnight at ±70 °C with intensifying screens. Standard human male DNA was run in the first 4 lanes (0.1–2.5 μg). The numbers 1–8 represent DNA samples obtained from a set of swabs related to a particular case. The letters above each lane indicate the anatomical source of the swab; v, vaginal; s, skin; r, rectal; bv, bloody vaginal; o, oral. The arrowhead denotes the position of the male specific band.

AMPLIFICATION OF VNTR LOCUS BY THE POLYMERASE CHAIN REACTION (PCR)

K. Kasai, Y. Nakamura and R. L. White

University of Utah
Salt Lake City, Utah

K. Kasai

National Research Institute of Police Science
Tokyo, Japan

Variable number of tandem repeat (VNTR) DNA markers are so hypervariable that they are very useful for paternity testing as well as individual identification. However, a usual Southern blotting method requires several days, radioactive materials and a relatively large amount of DNA sample. The polymerase chain reaction (PCR) method suggested the possibility of genotypic analysis of VNTR polymorphisms in shorter time and a very small amount of DNA. Also, the analysis would be free of any radioactive materials.

As the efficiency of DNA amplification is dependant on the size of DNA, we have chosen the VNTR locus (pMCT118) in which the repeating unit size is relatively small. By the results of DNA sequences, pMCT118 has shown more than 10 alleles within the range of 500 base pairs with 78% heterozygosity among 100 Caucasians. The pMCT118 locus has been shown to have a 16-base sequence as its repeating unit.

A repeating region was amplified by slightly modifying the PCR method reported by Saiki *et al.* (1988). The 28- and 29-based oligonucleotides derived from the flanking sequences of pMCT118 were synthesized for PCR reaction primers. Two ng each of human DNA was amplified in 25 ml of 67 mM Tris-HCl, pH 8.3, 6.7 mM MgCl2, 16.6 mM (NH4) 2So4, 10 mM 2-mercaptoethanol, 170 mμ/ml bovine serum albumin, 10% DMSO, 2.5 mM ATP, 2.5 mM CTP, 2.5 mM GTP, 2.5 mM TTP, 1.25 units of *Taq* polymerase and 2 mM each primer. After a denaturation at 95° C for 1.5 min., an annealing was done at 65° C for 1 min. and followed by an extension at 70° C for 10 min. (35 cycles).

Polymorphic bands were detectable by an ethidium bromide staining after an acrylamide gel electrophoresis. The difference of 1 repeating unit (16 bp) has been distinguished very clearly. More than 35 cycles of an amplification produced extra bands, and more than 10 ng of a template resulted in ladder bands because of the primers' shortness. Two ng of a template DNA and 30 to 35 cycles amplification gave the best result. It is important that high concentration (2 mM) of 28 and/or 29 base long primers make the intensities of bands almost identical regardless of their sizes.

We examined DNAs isolated from 50 unrelated Caucasian and Japanese. The distribution of alleles was between 384 bp and 624 bp, as expected for 15 to 30 repeat units.

These data supported that pMCT118 will be useful in forensic science study.

REFERENCES

Nakamura, Y., Carlson, M., Krapcho, K. and White, R. (1988). Isolation and mapping of a polymorphic DNA sequence (pMCT118) on chromosome 1P (D1S80), Nucleic Acid Res. 16:9364.

Saiki, R. L., Gelfand, D. H., Stoffel, S., Scharg, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase, Science 239:487–491.

BLANK PAGE

PCR INHIBITION AND BLOODSTAINS

S. Walsh and R. Higuchi

Cetus Corporation
Emeryville, California

E. Blake

Forensic Science Associates
Richmond, California

When extracted from blood stains, DNA is sometimes refractory to amplification by polymerase chain reaction. To investigate this phenomenon further, various RBC components were tested for their inhibitory effects on PCR.

MATERIALS AND METHODS

One method of obviating hematin inhibition in bloodstain samples is to simply add a very small fraction of the DNA isolate to the PCR so that the final hematin concentration stays below 0.8 uM. Methods for more efficient hematin removal from bloodstain extracts have also developed. It has been observed that hematin will not always pass through the membrane of the Centricon 30 (Amicon) ultrafiltration devices typically used to purify DNA for PCR. Some of the hematin remains in the retentate along with the DNA. This condition is perhaps due to aggregation of the hematin, which is very insoluble in aqueous solutions at neutral pH. It is also possible that binding of the hematin to the DNA contributes to the copurification phenomenon (Dougherty 1988), although in the absence of DNA, hematin is still retained. Hematin can be made to pass more freely through the membrane of a Centricon device by using an alkaline wash solution, which both solubilizes the hematin and potentially disrupts any hematin-DNA interactions. The use of an ultrafiltration device with a larger pore size (Centricon 100, Amicon) also helps in the removal of hematin.

Hematin inhibition experiments were performed with purified hematin from bovine blood (Sigma, St. Louis, MO). Purified protoporphyrin IX is also from Sigma. The PCR reactions are *Taq* DNA polymerase (PECI, Norwalk, CT) mediated reactions amplifying the DQα gene from human DNA using primers GH26 and GH27 (Scharf *et al.* 1986). A standard DQα PCR mixture contains the following in 100 μl: 50 mM KCl, 10 mM Tris (pH 8.3), 2.5 mM $M\theta$ C12, 188 uM each dATP, dCTP, dGTP and dTTP, 0.15 uM each biotinylated primer (GH26 and GH27) and 2.5 units *Taq* DNA polymerase (PECI). The amplification reactions were performed in a programmable heat block (DNA Thermal Cycler, PECI) set to heat at 94 °C for 30 sec. (denature), cool at 55 °C for 30 sec. (anneal), and incubate at 72 °C for 30 sec. (extend). After 40 cycles, the samples were incubated an additional 10 min. at 72 °C. Centricon 30 and Centricon 100 (Amicon, Danvers, MA) ultrafiltration devices are used as follows: DNA extract is added to a Centricon device containing 1.5 ml of Tris EDTA (10mM Tris, 0.1mM EDTA) or 1.5 ml of alkaline wash (0.4 M NaOH), then the devices are centrifuged in a fixed angle rotor for 20 min. at 5000 rpm. Samples are then washed with 2 ml of Tris EDTA and centrifuged as before for a total of at least two Tris EDTA washes. The PCR Amplified samples of PCR are analyzed on an ethidium bromide stained 3% NuSieve/1% agarose (Seakem GTG) gel run at about 10 V/cm for 30 min. Tank and gel buffer is 89 mM Tris, 89 mM Boric Acid, 2.5 mM EDTA, pH 8.3.

RESULTS

Hemoglobin, protoporphryin IX, FeC13 and FeSO4 inhibit PCR at concentrations as low as 20 μM (Table 1), and hematin is inhibitory at 0.8 μM (Table 2). The inhibition by hematin is of particular interest because of its abundance in aged bloodstains, because it partially co-purifies with DNA using standard extraction protocols. Plus it is such a potent inhibitor of PCR. Experiments with purified hematin have shown that hematin is not inhibitory at 0.4 μM, but that even large quantities of starting DNA (2 μg) would not amplify at a final hematin concentration of 0.8 μM. Hematin could be overcome by adding additional Taq polymerase (6.25 total units) or additional quantities of primers to the PCR (an additional 15 pmoles of each primer). These results suggest that hematin is a competitive inhibitor of Taq polymerase.

Table 1. INHIBITION OF PCR BY RED BLOOD CELL COMPONENTS

Inhibitor Conc. (μM)	Inhibitors						
	glo-bin*	hemo-globin	FeCl(3)	FeSO(4)	he-matin	por-phy-rin	none
20	0	4	2	5	0	0	10
40	0	3	1	–	0	–	
80	0	2	0	0	0	–	

0 = no PCR product
1-10 = range of PCR product yield
*visible precipitate formed upon addition of this regent.

REFERENCES

Dougherty, G. (1988). Intercalation of tetracationic metalloporphyrins and related compounds into DNA, J. Inorg. Biochem. 34:95–103.

Scharf, S. J., Horn, G. T. and Erlich, H. A. (1986). Direct cloning and sequence analysis of enzymatically amplified genomic sequences, Science 88:1076–1078.

Table 2. SUMMARY OF HEMATIN INHIBITION DATA

He-matin Conc. (uM)	Quantity of Starting DNA (ng)					
	5ng.	10	20–25	40–50	500	2000
20	+	+	+	+		
2	+	+	+	+		
1.6	+	+	+	+		
1.2	+	+	+	+		
0.8	+	+	+	+	+	+
0.4	–	–	–	–	–	–
0.2	–	–	–	–		
0.02	–	–	–	–		
0.002	–	–	–	–		

+ = inhibition (no visible PCR product)

SECTION III

PANEL DISCUSSIONS

BLANK PAGE

PANEL DISCUSSION
DNA—INTERNATIONAL EXPERIENCES

MODERATOR: Bruce Budowle, FBI Laboratory

PANEL MEMBERS: Brian Parkin, Metropolitan Police Forensic Science Laboratory
Hermann Schmitter, Bundeskriminalamt
Gary Shutler, Royal Canadian Mounted Police
David Werrett, Home Office Forensic Science Service

The purpose of this panel was to discuss the experiences obtained from the panel members, many of whom have been involved in the original DNA research for forensic applications. One of the important considerations in this topic area, is that DNA is relatively new (last 5 years) in the forensic field and many of the experiences have been learned during the research and case work implementation stages.

Each panel member had previously discussed their experiences with DNA in earlier lectures so the main dialogue was responding to questions from the audience.

One concern was that some individuals have been reluctant to initiate DNA analysis in their laboratory because although there is one type of technology available right now there is something new coming along in 6 months; therefore, maybe it would be better to wait. There could be considerable expenses in buying new equipment that might be obsolete very shortly and the laboratory would have the added expense of buying all new equipment. In response the consensus was that in rapidly evolving technology, such as DNA, for the foreseeable future, one will be in that position. Therefore, it would be preferable to begin with an established technique and gain experience with molecular biology technology issues concerning evidentiary material and interpretations, then, as a new technique becomes available, it would be more easily implemented into the laboratory protocol. This approach was supported by at least two of the laboratories represented on the panel, the CRE and the FBI, both of which implemented testing using RFLP and are presently investigating PCR technology.

There was then some discussion directed towards the panel members regarding if they had initiated country wide data bases or indexes of DNA. The response was that these laboratories were still in the discussion stage for developing policy as to how a DNA indices might be implemented. Again, as of this date there have not been sufficient studies to determine the feasibility of a national or international database(s).

The discussion then turned to the policy of whether or not to analyze cases where there are no suspects. The general response was that given the large case loads it is currently not possible to work this type of case. Eventually, the DNA profiles from all cases with no suspects could be analyzed and stored in a national DNA data bank for comparison with known individuals.

One question was on the use of ethidium bromide and the effects on migration. One response was that the molecular biological sciences had stopped using this compound in the previous decade and recommended against adding this to samples. In contrast, other laboratories use it routinely. Additionally, recognizing that forensic samples are not in pristine condition, it would be advantageous to know how much DNA has been extracted for analysis.

The consensus of the panel members was that considering the type of specimens that are available for forensic analysis it would be extremely important to obtain as much information as possible about the case and about the samples. Finally, the advice was to expect anything and be prepared for anything.

BLANK PAGE

PANEL DISCUSSION
DNA - NATIONAL EXPERIENCES

MODERATOR: John Waye, Royal Canadian Mounted Police

PANEL MEMBERS: Dwight Adams, FBI Laboratory
William Eubanks, FBI Laboratory
Paul Ferrara, Virginia Bureau of Forensic Sciences
Willard Stuver, Metro-Dade Police Department

This panel discussed the experiences obtained from the panel members who have been involved in the implementation of DNA into case work in forensic laboratories. The panel members pointed out that since few laboratories are operational for the analysis of DNA, experiences in this field are somewhat limited. Thus, most of the experiences have been in the implementation of DNA into case work.

The panel members had previously discussed their experiences with implementing DNA in case work in earlier lectures so the panel members primarily responded to questions from the audience. The laboratories represented on the panel have equipped and trained their respective laboratories to have at least two examiners and one technician perform DNA analysis.

A question was raised as to whether submitters would need to change anything in their sampling protocol since DNA analysis is now available for sexual assault cases. The response was yes, for one thing, more swabs of both the vaginal and cervical areas would need to be collected since additional tests are now being conducted. Also, the preferred sample type would be vaginal swabs and not vaginal aspirates, as the liquid specimen, if not properly stored, would allow degradation, whereas the dried swab would not promote degradation. Regarding bloodstains, air dried stains are the best way to preserve this type of specimen. Wet bloodstains should not be submitted. Regarding storage, all specimens, extracts and membranes should be stored frozen.

A question addressed to the panel was if running DNA cases saved any steps, that is, do you still determine if the stain is human in origin and determine if it is blood and/or semen. The response is always perform a typical serological examinations first.

The panel members and audience then discussed the fact that some states now routinely collect blood samples from convicted sex offenders. The blood samples are provided to the state forensic laboratory. It is recommended that the DNA be extracted and stored as duplicates and blood should also be stored as duplicates. Thus, one would have duplicate blood specimens and DNA extracts stored for future analysis.

The next question was on reporting procedures, specifically how laboratories that are performing DNA analysis are reporting out their results. The FBI puts out a simple report that states analyses of DNA were performed on known and unknown specimens. If there is no match the report will indicate this. If there is a match, the report will state that the samples match and provide a probability. In addition, the probes used for matches will be reported. If probe results are not found, the report will indicate failure to obtain a probe result could be due to insufficient DNA or degraded DNA.

The last question to the panel sought to determine if there were any instances where DNA did not work but conventional serology (enzymes) did work? As of this date, there actually have been good results from DNA where there was limited success with conventional systems.

BLANK PAGE

PARTICIPANTS

Ronald T. Acton
University of Alabama
P.O. Box 201 - Volker Hall
Birmingham, Alabama 35294

Dwight E. Adams
Laboratory Division
Federal Bureau of Investigation
Washington, D.C. 20535

Ms. Haruka Akasaka
Scientific Criminal Laboratory
Metropolitan Police Department
2-1-1, Chiyoda-Ku-Kasumigaseki
Tokyo 100, Japan

Robert C. Allen
Department of Pathology
Medical University of South Carolina
171 Ashley Avenue
Charleston, South Carolina 29425

F. Samuel Baechtel
Laboratory Division
Federal Bureau of Investigation
Quantico, Virginia 22135

Bobby D. Blackburn
Laboratory Division
Federal Bureau of Investigation
Quantico, Virginia 22135

Edward T. Blake
Forensic Science Associates
3053 Research Drive
Richmond, California 94806

Frank Booth
Kansas City Police Department
1525 Holmes
Kansas City, Missouri 64108

Bruce Budowle
Laboratory Division
Federal Bureau of Investigation
Quantico, Virginia 22135

Martin R. Buoncristiani
Forensic Science Research Group
University of California
Berkeley, California 94701

C. Thomas Caskey
Institute for Molecular Genetics
Baylor College of Medicine
Houston, Texas 77008

Andrew Cawood
ICI Diagnostics
20271 Goldenrod Lane
Germantown, Maryland 20874

George W. Clarke
Office of the District Attorney
220 W. Broadway
San Diego, California 92101

Howard C. Coleman
Genelex Corporation
1000 Seneca Street
Seattle, Washington 98101

Catherine T. Comey
Laboratory Division
Federal Bureau of Investigation
Quantico, Virginia 22135

Raphael Coquoz
Institute of Forensic Science
University of Lausanne
place du Chateau 3
Lausanne, Switzerland 1005

Robin W. Cotton
Cellmark Diagnostics
P.O. Box 1000
Germantown, Maryland 20874

Henry Erlich
Cetus Corporation
1400 53rd Street
Emeryville, California 94608

William G. Eubanks
Laboratory Division
Federal Bureau of Investigation
Washington, D.C. 20535

Paul Ferrera
Bureau of Forensic Science
Commonwealth of Virginia
1 N. 14th Street
Richmond, Virginia 23219

Lawrence I. Grossman
Department of Molecular Biology & Genetics
Wayne State University
540 East Canfield Avenue
Detroit, Michigan 48201

Harry W. J. Harding
Forensic Science Centre
21 Divett Place
Adelaide, South Australia 5000

John Harding
Bethesda Research Laboratories
Life Technologies Incorporated
8717 Grovemont Circle
Gaithersburg, Maryland 20877

John W. Hicks
Laboratory Division
Federal Bureau of Investigation
Washington, D.C. 20535

Bruce H. Howard
Laboratory of Molecular Biology
National Institutes of Health
9000 Rockville Pike
Bethesda, Maryland 20892

Alec J. Jeffreys
Department of Genetics
University of Leicester
University Road
Leicester, LE1 1RH, England

Barbara L. Johnson
Los Angeles County Sheriff's Laboratory
2020 W. Beverly Boulevard
Los Angeles, California 90057

Donald M. Johnson
Technical Services Division
Federal Bureau of Investigation
Washington, D.C. 20535

M. Roger Kahn
Metro-Dade Police Crime Laboratory
1320 NW 14th Street
Miami, Florida 33125

Kentaro Kasai
National Research Institute of Police Science
6 Sanban-cho, Chiyoda-ku
Tokyo, 102, Japan

James J. Kearney
Laboratory Division
Federal Bureau of Investigation
Quantico, Virginia 22135

Kenneth C. Konzak
California Department of Justice
Forensic Science Research Group
University of California
Berkeley, California 94598

Terry A. Landers
Bethesda Research Laboratories
Life Technologies Incorporated
8717 Grovemont Circle
Gaithersburg, Maryland 20877

Margaret E. Lawton
Chemistry Division
Department of Scientific and Industrial Research
P.O. Box 2224
Auckland, New Zealand

Carl R. Merril
Clinical Neurogenetics Branch
National Institutes of Health
9000 Rockville Pike
Bethesda, Maryland 20892

Keith L. Monson
Laboratory Division
Federal Bureau of Investigation
Quantico, Virginia 22135

James L. Mudd
Laboratory Division
Federal Bureau of Investigation
Quantico, Virginia 22135

Pamela J. Newall
Centre of Forensic Sciences
25 Grosvenor Street
Toronto, Ontario, M7A 2G8 Canada

Elaine Pagliaro
Connecticut State Police Laboratory
294 Colony Street
Meriden, Connecticut 06450

Brian Parkin
Metropolitan Police Forensic Science Laboratory
109 Lambeth Road
London, SE1 7LP, England

Vincenzo Pascali
Forensic Medicine Institute
Universita' Cattolica Sacro Cuore
Largo Francesco Vito, 1
Rome, Italy 00168

Lawrence A. Presley
Laboratory Division
Federal Bureau of Investigation
Washington, D.C. 20535

Rebecca L. Reynolds
Forensic Science Research Group

University of California
Berkeley, California 94720

Antti Sajantila
Laboratory for Molecular Genetics
National Public Health Institute
Mannerheimintie 166
Helsinki, Finland 00300

Herman Schmitter
Kriminaltechnisches Institut
Bundeskriminalamt
Wiesbaden KT 24
Federal Republic of Germany

George F. Sensabaugh
Forensic Science Research Group
University of California
Berkeley, California 94720

Kathleen T. Sheridan
Cellmark Diagnostics
P.O. Box 1000
Germantown, Maryland 20874

Gary Shutler
Central Forensic Laboratory
Royal Canadian Mounted Police
P.O. Box 8885
Ottawa, Ontario, K1G 3M8 Canada

David A. Stoney
Department of Criminal Justice
University of Illinois
Chicago, Illinois 60680

Willard C. Stuver
Metro-Dade Police Crime Laboratory
1320 NW 14th Street
Miami, Florida 33125

Susan L. Swarner
Contra Costa Sheriff's Laboratory
1122 Escobar Street
Martinez, California 94553

Richard L. Tanton
Palm Beach County Sheriff's Department
3228 Gun Club Road
West Palm Beach, Florida 33406

Lowell Tilzer
Kansas University Medical Center
39th & Rainbow
Kansas City, Kansas 66103

Cecilia von Beroldingen
Forensic Science Research Group
University of California
Berkeley, California 94720

Patric S. Walsh
Cetus Corporation
1400 53rd Street
Emeryville, California 94608

John S. Waye
Central Forensic Laboratory
Royal Canadian Mounted Police
P.O. Box 8885
Ottawa, Ontario, K1F 3M8 Canada

David J. Werrett
Central Research Establishment
Home Office Forensic Science Service
Aldermaston, Reading
Berkshire, RG7 4PN, England

Raymond L. White
Howard Hughes Medical Institute
University of Utah
Salt Lake City, Utah 84132

BLANK PAGE

AUTHORS

BLANK PAGE

END

FILMED

03/16/92